





Db 156 ProSerSerProAlaProLeuProProAlaProAlaSerProProSerProAlaProPro 175  
QY 54 CCACGAGCTCTCAAACTCGGCTCTTGAGACTCCGCGATCCCGCGCGCGC 1  
Db 176 ProProAlaProProProAlaPro-----HisProProSerProPro 189  
RESULT 3  
ADU02754  
ID ADU02754 standard; protein; 483 AA.  
XX ADU02754;  
AC ADU02754;  
XX 27-JAN-2005 (first entry)  
XX DE Novel human polypeptide seqid 1221.  
XX cytosstatic; antipsoiatric; antiinflammatory; gene therapy; Nanodisc;  
XX proliferative disorder; inflammatory disorder; immune disorder;  
XX metabolic disorder; bone disorder; CNS disorder; cancer; psoriasis;  
XX ulcerative colitis; human.  
XX Homo sapiens.  
XX WO2004093804-A2.  
XX 04-NOV-2004.  
XX 19-APR-2004; 2004WO-US012047.  
XX 18-APR-2003; 2003US-0463708P.  
XX 18-APR-2003; 2003US-0463732P.  
XX 02-MAY-2003; 2003US-0467199P.  
XX 02-MAY-2003; 2003US-0467230P.  
XX 19-MAY-2003; 2003US-0471306P.  
XX 19-MAY-2003; 2003US-0471336P.  
XX 08-JUL-2003; 2003US-0485223P.  
XX 08-JUL-2003; 2003US-0485224P.  
XX 14-JUL-2003; 2003US-0486446P.  
XX 14-JUL-2003; 2003US-0486480P.  
XX 08-AUG-2003; 2003US-0493577P.  
XX 08-AUG-2003; 2003US-0493577P.  
XX 08-SEP-2003; 2003US-0505059P.  
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.  
XX Lee E, Hestir K, Chu K, Masuoka L, Williams LT;  
XX WPI; 2004-775861/76.  
XX N-PSDB; ADU02022.  
XX New first nucleic acid molecule comprising a polynucleotide sequence  
XX given in the specification, useful in preparing a composition for  
XX diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.  
XX Claim 14; SEQ ID NO 1221; 291pp; English.  
XX  
XX The invention describes a new first nucleic acid molecule comprising a  
XX polynucleotide sequence given in the specification. Also described are:  
XX an animal injected with the nucleic acid molecule; a second nucleic acid  
XX molecule comprising a second polynucleotide sequence that is at least  
XX about 70, 80, 90 or 95% homologous to the first nucleic acid molecule or  
XX that hybridises to the first polynucleotide sequence under high  
XX stringency conditions; a vector comprising the nucleic acid molecule and  
XX a promoter that drives the expression of the nucleic acid molecule; a  
XX host cell transformed, transfected, transduced or infected with the  
XX nucleic acid molecule; a nucleic acid composition comprising a carrier or  
XX a buffer and one or more compositions comprising the nucleic acid  
XX molecule, vector or host cell; a substantially purified polypeptide; an  
XX animal injected with the polypeptide; a polypeptide composition  
XX comprising the polypeptide molecule and a carrier or buffer; a cell  
XX culture medium comprising the polypeptide or transfected cells  
XX transfected with the polynucleotide; making a transformed, transfected,  
XX transduced, or infected host cell; synthesising Nanodiscs simultaneously

CC and for synthesising a series of simultaneously-synthesised Nanodiscs  
CC sequentially utilising a dynamic system; preparing a hydrophobic protein  
CC for determination of crystal structure; immunising a non-human animal;  
CC screening for modulators of hydrophobic protein activity; a diagnostic  
CC kit; determining the presence of the nucleic acid molecule or its  
CC complement; determining the presence of an antibody to the polypeptide in  
CC a sample; an antibody specifically recognising, binding to or modulating  
CC the biological activity of at least one polypeptide encoded by a nucleic  
CC acid molecule or its biologically active fragment; an antibody  
CC composition comprising the antibody and a carrier; a bacteriophage, where  
CC the antibody is displayed on the bacteriophage; a bacterial cell  
CC comprising the bacteriophage; a non-human animal injected with the  
CC antibody composition; a host cell that secretes the antibody; making an  
CC antibody; diagnosing a disease, disorder, syndrome, or condition  
CC comprising cancer, or proliferative, inflammatory, immune, metabolic,  
CC bone, CNS, genetic, bacterial and viral diseases, disorders, syndromes or  
CC conditions in a patient; a modulator composition comprising a modulator  
CC and a carrier; gene therapy; prophylactic or therapeutic treatment of a  
CC subject; an isolated modified cell comprising at least one first  
CC heterologous nucleic acid molecule, where the first heterologous nucleic  
CC acid molecule comprises a first polynucleotide sequence that encodes a  
CC first polypeptide; a non-human animal deficient in the polypeptide or  
CC that over-expresses the polypeptide; isolated tissues derived from the  
CC non-human animal; and one or more cells derived from the non-human  
CC animal. The nucleic acid is useful in preparing a composition for  
CC diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.  
CC This is the amino acid sequence of a novel human polypeptide of the  
CC invention.  
XX  
XX SQ Sequence 483 AA;  
XX  
XX Alignment Scores:  
XX Pred. No.: 7.49e-05 Length: 483  
XX Score: 142.50 Matches: 39  
XX Percent Similarity: 46.1% Conservative: 8  
XX Best Local Similarity: 38.2% Mismatches: 32  
XX Query Match: 15.4% Indels: 23  
XX DB: 8 Gaps: 4  
XX  
XX US-10-071-510A-16 (1-493) x ADU02754 (1-483)  
QY 273 CCATTTCTCCCGAATCCACTTCCTCGACGATGCGAGTTTGTATGAGGCTCA 214  
Db 240 ProSerProSerProGlnProLeuProSerSerLeuProPro-----Ser 254  
QY 213 GCAGACTTCTTCACAGACCTTCAGGCCCGCACGACTCTGCGAGCTCCCTGAGCTCT 154  
Db 255 SerProSerProProProProSerProProSerSerSerLeuPro-----Pro 270  
QY 153 CCGTCAACACGCGACGACCTCCGAGAAAGACTTCCATCAGACGACCGCCCTGGGCTT 94  
Db 271 ProAlaProProAlaAlaProProProSerLeuProSerProProSerProPro 290  
QY 93 CCACGAGGAGACGCT-----GGGCTCTCTTCT 67  
Db 291 ProProSerSerSerProProProProLeuSerAlaProProProProSerProProSer 310  
QY 66 CCGGGAATTTCTGCGACGACGCTCTCAAACTCGGCTCTTGAGACTCGGATCCCGCGGC 7  
Db 311 ProSerProSerProProAlaPro-----ProLeuSerProSerProProPro 327  
QY 6 CCGCGC 1  
Db 328 ProPro 329  
RESULT 4  
ADT60739  
ID ADT60739 standard; protein; 238 AA.  
XX AC ADT60739;  
XX 13-JAN-2005 (first entry)  
XX

DE	Plant polypeptide, SEQ ID 10816.
XX	
XX	Plant; transgenic; cold tolerance; growth rate; drought tolerance;
KM	disease resistance; galactomannan production; plant growth regulator;
KM	heat tolerance; herbicide tolerance; lignin production;
XX	extreme osmotic condition tolerance; pathogens resistance;
KM	pest resistance; yield improvement; seed oil yield; seed protein yield.
XX	
OS	Viridiplantae.
XX	
PN	US2004216190-A1.
XX	
PD	28-OCT-2004.
XX	
PF	18-DEC-2003; 2003US-00739930.
XX	
PR	28-APR-2003; 2003US-00424599.
XX	
PR	28-APR-2003; 2003US-00425115.
XX	
PA	(KOVA/) KOVALIC D. K.
XX	
P1	Kovalic DK;
XX	
DR	WPI: 2004-757369/74.
XX	
PT	New recombinant DNA constructs useful in the field of biochemistry and
XX	genetics, and in particular for producing transgenic plants with improved
PT	biological characteristics.
XX	
PS	Claim 2; SEQ ID NO 10816; 14pp; English.
XX	
CC	The invention relates a recombinant DNA construct comprising a
CC	polynucleotide having any of 5544 nucleotide sequences (CDNAs SEQ ID NO:
CC	1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
CC	(SEQ ID NO: 5545-11088) . The CDNAs and proteins are from corn, soybean,
CC	Arabidopsis, wheat and rape but the specification does not indicate which
CC	sequences is derived from which organism. Also included is a method of
CC	producing a plant having an improved property, comprising transforming a
CC	plant with a recombinant DNA construct comprising a promoter region
CC	functional in a plant cell operably joined to a polynucleotide encoding a
CC	polypeptide associated with the property, and growing the transformed
CC	plant. The property is selected from improving plant cold tolerance, for
CC	manipulating growth rate in plant cells by modification of the cell cycle
CC	pathway, for improving plant drought tolerance, for providing increased
CC	resistance to plant disease, for galactomannan production, for production
CC	of plant growth regulators, for improving plant heat tolerance, for
CC	improving plant tolerance to herbicides, for increasing the rate of
CC	homologous recombination in plants, for lignin production, for improving
CC	plant tolerance to extreme osmotic conditions, for improving plant
CC	tolerance to pathogens or pests, for yield improvement by modification of
CC	photosynthesis, for modifying seed oil yield and/or content, for
CC	modifying seed protein yield and/or content, for yield improvement by
CC	modification of carbohydrate, nitrogen or phosphorus use and/or uptake
CC	and for yield improvement by providing improved plant growth and
CC	development under at least one stress condition. The polynucleotide may
CC	also encode a plant transcription factor. The methods and compositions of
CC	the present invention are useful in the field of biochemistry and
CC	genetics, in particular for producing transgenic plants with improved
CC	biological characteristics such as increased yield, improved nitrogen
CC	flow, increasing plant tolerance to cold or heat, improving plant
CC	tolerance to extreme osmotic and drought conditions, and improving plant
CC	tolerance to plant pests or pathogens. They can also be used in physical
CC	arrays of molecules, plant breeding markers, computer-based storage and
CC	analysis systems. The present sequence is one of the 5544 plant protein
CC	sequences of the invention. Note: The sequence data for this patent did
CC	not form part of the printed specification, but was obtained in
CC	electronic format directly from USPTO at
CC	seqdata.uspto.gov/sequence.html?DocID=20040216190.
XX	
XX	Sequence 238 AA;
SO	
Alignment Scores:	
Preld. No.:	8.3e-05 length: 238

Score:	141.00	Matches:	39
Percent Similarity:	42.6%	Conservative:	7
Best Local Similarity:	36.1%	Mismatches:	40
Query Match:	15.3%	Indels:	22
DB:	8	Gaps:	4
US-10-071-510A-16 (1-493) x ADT60739 (1-238)			
QY	273 CCATTTTCTTCCCGAATTCACCTTCATCTCTGACGATGCCAGTTTCTGATGAGGCTCA	214	
DB	56 ProAlaGlyAsnProProProAlaThrProSerAlaProProAlaProThrThrProProGln	75	
QY	213 CGAGACTTCTTCCGACGAGCC-----TCAAGGCCGCGCACGACTGCGCA	169	
DB	76 AlaProGlnAlaProAlaThrProProProAlaProGlnAlaProAlaThrProProPro	95	
QY	168 GCTCCCTGAGCTCCTCTGCA-----	148	
DB	96 AlaPro***ThrProProAlaThrProProProAlaProThrThrProProProAlaPro	115	
QY	147 CCACGGCAGCACCTTCGAGAGACTTCTCATCACGACGAGCCCTGCGCTTCACCA	88	
DB	116 ProThrGlnProProProAlaProThrThrProProProSerProProAlaThrProPro	135	
QY	87 GGGACAGCTGGG-----CTGCTTCTCCGGGAATTTGCGCACGAGCCCTCAAACTCG	34	
DB	136 ProAlaProAlaThrProProProProSerProProMetSerProProProAlaThrProPro	155	
QY	33 CCTCTTGGGACTCGGCAATCCCG 10		
DB	156 ProMetAlaThr-----ProPro 161		
RESULT 5			
ADN21309			
ID	ADN21309 standard; protein; 639 AA.		
XX	ADN21309;		
AC	ADN21309;		
DT	02-DEC-2004 (first entry)		
XX			
DE	Bacterial polypeptide #362.		
XX			
KM	Recombinant DNA construct; transformed plant; improved plant property;		
KM	cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;		
KM	pathogen tolerance; pest tolerance; plant disease resistance;		
KM	cell cycle pathway modification; plant growth regulator;		
KM	homologous recombination; seed oil yield; protein yield; carbohydrate;		
KM	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;		
KM	bacterial polypeptide.		
XX			
OS	Bacteria.		
XX			
PN	US2003233675-A1.		
XX			
PD	18-DEC-2003.		
XX			
PF	20-FEB-2003; 2003US-00369493.		
XX			
PR	21-FEB-2002; 2002US-0360039P.		
XX			
PA	(CAOY/) CAO Y.		
PA	(HINK/) HINKLE G J.		
PA	(SLAT/) SLATER S C.		
PA	(CHEN/) CHEN X.		
PA	(GOLD/) GOLDMAN B S.		
XX			
PI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;		
XX			
DR	WPI; 2004-061375/06.		
XX			
PT	New recombinant DNA construct comprising a promoter positioned to provide		
PT	for expression of a polynucleotide encoding a polypeptide from a		
PT	microbial source, useful for producing plants with improved properties.		



XX Claim 1; SEQ ID NO 3962; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plants with  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polypeptide used in the  
CC scope of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).

XX Sequence 639 AA;

Alignment Scores:  
Pred. No.: 0.000229 Length: 639  
Score: 138.00 Matches: 39  
Percent Similarity: 43.8% Conservative: 3  
Best Local Similarity: 40.6% Mismatches: 36  
Query Match: 15.0% Indels: 18  
DB: Gaps: 4

US-10-071-510A-16 (1-493) x ADN21309 (1-639)

QY 255 CCACCTTCACCTCTCTGACATGCCAGTTCGTGATGAGCTCAGACACTTCTT---CCA 199  
DB 423 ProLeuProProLeuProAlaPro-----GlyProAlaProProLeuProPro 437  
QY 198 GCAGCTTCAGAGCGCGCAGACTCTGCGAGCTCCTGAGCTCTCTGACCAACGCGCAG 139  
DB 438 AlaSerSerArgProProProMetLeuProThr-----ArgSerProAlaProProGln 455  
QY 138 CACCTTCGCGAGAGAACTTCTCCATCAGCAGCGAGCCCTGCGCTTCCA-----91  
DB 456 AlaProProLeuProThrSerArgAlaProProProProProLeuProAlaThrGlnAla 475  
QY 90 -----CCAGGACAGCTGGGCTCTCTCTCCGGAATTCGCCACCA 49  
DB 476 ProProProProProLeuProAlaThrSerAlaProProProProProProAlaProPro 495  
QY 48 GCCTTCACAACTGGCTCTTGAGACTCGGATCCCCGGGCGCGC 1  
DB 496 AlaProProAlaProProLeuProAlaAlaAlaAlaProProProPro 511

RESULT 6  
AAB74209  
AAB74209 standard; protein; 731 AA.  
AC AAB74209;  
DT 17-MAY-2001 (first entry)  
XX  
XX Protein encoded by Arabidopsis gene #4.  
XX  
XX Proline-rich extensin-like receptor kinase; PERK; resistance; plant.  
XX  
XX Arabidopsis thaliana.  
OS

XX .MO200114563-A1.

XX 01-MAR-2001.

XX 18-AUG-2000; 2000WO-CA000966.

XX 19-AUG-1999; 99US-0149466P.

XX 13-OCT-1999; 99US-0159122P.

XX (GORI/) GORING D.

XX (SILV/) SILVA N.

XX Goring D, Silva N;

XX WPI; 2001-244305/25.

PT New proline-rich, extensin-like receptor kinase nucleic acids and  
PT polypeptides useful for increasing plant wounding or pathogen resistance,  
PT or for producing transgenic plants with increased wounding or pathogen  
PT resistance.

XX Example; Fig 14; 91pp; English.

XX The present invention relates to proline-rich extensin-like receptor  
CC kinase (PERK). The PERK nucleic acids and polypeptides are useful for  
CC increasing the resistance of plants to wounding and pathogens. These are  
CC also useful for producing transgenic plants with increased wounding and  
CC pathogen resistance compared with a wild type plant, as well as in assays  
CC for identifying and developing compounds to inhibit and/or enhance  
CC polypeptide function directly

XX Sequence 731 AA;

Alignment Scores:  
Pred. No.: 0.000378 Length: 731  
Score: 136.00 Matches: 52  
Percent Similarity: 33.5% Conservative: 8  
Best Local Similarity: 29.1% Mismatches: 54  
Query Match: 14.7% Indels: 65  
DB: Gaps: 6

US-10-071-510A-16 (1-493) x AAB74209 (1-731)

QY 442 GCCCCAGACAGAGAGGCGCTGTGCTGAGACCCCTACCC-----AC 401  
DB 10 AlaProAlaThrSerProProAlaMetSerLeuPro-ProAlaApsSerValProAspTh 29  
QY 400 CTCTTGACCTGGGGCGCTGCGCTCTCCACAGCCCTGCTAGACAGACTCAGCGCTCGA 341  
DB 29 rSerSerProProAlaProProLeuSerProLeuPro-----41  
QY 340 CGATGCTGGGAATAGATCATGAGATTGATGAGAAATCTGACTTTGGATGTTGTG 281  
DB 41 -----41  
QY 280 GTGMAAACATTTCTTCCCGAATCCACTTCATCTCTGAGATGCCAGTTTGTGATG 221  
DB 42 ---ProProLeuSerSerProProProProProProProProProProLeuSerAlaProTh 60  
QY 220 AGGCTCAGACAGACTTCTTCCAGACCTCAAGCGCGCAGAGACTCTGCGAGCTCC-- 163  
DB 60 rAlaSerProProProLeuProAlaGlnSerProProSerProProIleGlnSerProPr 80  
QY 162 -----TGAGCTCTCTGCTGACACCGCAGCACACCT-----133  
DB 80 oProProLeuLeuGlnSerProProProProProProLeuGlnSerProSerProSerPr 100  
QY 132 -----CCGAGAAAGCTTCTCC 116  
DB 100 oHisValSerAlaProSerGlySerProProLeuProPhleuProAlaLysProSerPr 120  
QY 115 ATCACACAGCAGGCCCT-----GCGCTTCCACAGGGAAGAGCTGGGCTCTCTCC 65

```
Db 120 oProProSerSerProProSerGluThrValProProGlyAsnThrIleSerProProPr 140
Oy 64 GGAATTTCGCCACCAAGCCTCTCAAACTGGCCTCTTGAGACTCGGACATCCCGG 10
Db 140 oArgSerLeuProSerGluSer---ThrProProValAsnThrAlaSerProPro 157

RESULT 7
ABR82942
ID ABR82942 standard; protein, 731 AA.
AC ABR82942;
XX
XX 31-MAY-2002 (first entry)
DE Herbicidally active polypeptide SEQ ID NO 2413.
XX
XX Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
XX WO200210210-A2.
XX
XX 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP009892.
XX
XX 28-AUG-2001; 2001WO-EP009892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidner M;
XX
XX WPI: 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms.
XX
XX Claim 5; SEQ ID NO 2413; 261bp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins (ABR82942)
XX for herbicidally active compounds, comprising aligning and comparing
XX nucleic acid or amino acid sequences from plant with nucleic acid or
XX amino acid sequences from non-plant organisms using suitable search
XX parameters, where plant sequences having an E-value greater by a factor
XX of 3 than the E-value of most similar non-plant sequences are selected.
XX The polypeptides or nucleic acids encoding them are useful for
XX identifying modulators. The identified modulators are useful as
XX herbicides
XX
XX SQ Sequence 731 AA;
XX
XX Alignment Scores:
XX Pred. No.: 0.000378 Length: 731
XX Score: 136.00 Matches: 52
XX Percent Similarity: 33.5% Conservative: 8
XX Best Local Similarity: 29.1% Mismatches: 54
XX Query Match: 14.7% Indels: 65
XX DB: 5 Gaps: 6
XX
XX US-10-071-510A-16 (1-493) x ABR82942 (1-731)
XX
XX 442 GCCCAGACAGGAGGAGGCGCTGTGCTGAGACCCCTACCC-----AC 401
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 10 AlaProAlaThrSerProProAlaMetSerLeuPro-ProAlaAspSerValProAspTh 29
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 400 CTCTTGACTGGGGCGCTGCGCTTCTCCACAGCCCTGCTAGACAGACTACGGGTGA 341
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 29 rSerSerProProAlaProProLeuSerSerLeuPro----- 41
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 340 CGATCCTCGGAATAGATCCATGGATTGATGAGAAATCTGACTTGGAATGTTGTTG 281
```

```
Db 41 ----- 41
Oy 280 GTGAACCATTTTCTTCCCGGAATCCACTTCATCTCTGACAGATGCCAGTTTCTGATG 221
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 42 ---ProProLeuSerSerProProProLeuProSerProProLeuSerAlaProTh 60
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 220 AGGCTCAGCAGACTTTCTTCCAGCAGCCTCAAGGCCGCCACAGACTTGACAGCTCC-- 163
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 60 rAlaSerProProProAlaProValGluSerProProSerProProIleGluSerProPr 80
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 162 -----TGAGCTCTCTGACCAAGCAGCAGCCT----- 133
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 80 oProProLeuLeuGluSerProProProProProLeuGluSerProSerProSerPr 100
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 132 -----CCGAGAGACTTCTCC 116
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 100 ohIeValSerAlaProSerGlySerProProLeuProProLeuProAlaProSerPr 120
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 115 ATCACCAGCAGCCT-----GCGCTTCCACAGGAGCAGCTGCGCTCTCTCC 65
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 120 oProProSerSerProProSerGluThrValProProGlyAsnThrIleSerProProPr 140
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 64 GGAATTTCGCCACCAAGCCTCTCAAACTGGCCTCTTGAGACTCGGACATCCCGG 10
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 140 oArgSerLeuProSerGluSer---ThrProProValAsnThrAlaSerProPro 157

RESULT 8
ABR82942
ID ABR82942 standard; protein, 731 AA.
XX
XX ABR82942;
XX
XX 18-DEC-2003 (first entry)
XX
XX Arabidopsis PERK1 receptor related protein.
XX
XX Arabidopsis PERK1 receptor related protein.
XX
XX PERK1; transgenic; plant; proline-rich extensin-like receptor kinase;
XX wound; pathogen resistance; plant growth; seed production.
XX
XX Arabidopsis thaliana.
XX
XX WO2003072763-A1.
XX
XX 04-SEP-2003.
XX
XX 28-FEB-2003; 2003WO-CA000274.
XX
XX 28-FEB-2002; 2002CA-02373903.
XX
XX 28-FEB-2002; 2002US-00084644.
XX
XX (GORI/) GORING D.
XX (SILV/) SILVA N.
XX (HAEF/) HAEFANI Y Z.
XX
XX Goring D, Silva N, Haeffani YZ;
XX
XX WPI: 2003-712727/67.
XX N-P8DB; ACP36553.
XX
XX Producing a transgenic plant having an increased plant resistance, plant
XX growth or seed production comprises transforming a plant with a nucleic
XX acid molecule having a Proline-rich Extensin-like Receptor Kinase
XX activity.
XX
XX Disclosure; Fig 14; 123bp; English.
XX
XX The invention relates to producing a transgenic plant having increased
XX plant height, number of branches, number of seed pods and/or seed
XX production compared to a non-transgenic plant, and/or quicker flowering
XX or later senescence compared to a non-transgenic plant. The method
XX involves transforming a plant with a vector including a Proline-rich
XX Extensin-like Receptor Kinase (PERK) nucleic acid molecule or a nucleic
```



Db	84_ProProSerProProSerProSer-----ProLeuSerProThrProProPro 99			
RESULT 10				
ABG14000				
ID	ABG14000	standard; protein; 598 AA.		
XX				
AC	ABG14000;			
XX				
DT	18-FEB-2002	(first entry)		
XX				
DE	Novel human diagnostic protein #13991.			
XX				
KM	Human; chromosome mapping; gene mapping; gene therapy; forensic;			
XX	food supplement; medical imaging; diagnostic; genetic disorder.			
OS	Homo sapiens.			
XX				
PN	MO200175067-A2.			
XX				
PD	11-OCT-2001.			
XX				
PF	30-MAR-2001; 2001MO-US008631.			
XX				
PR	31-MAR-2000; 2000US-00540217.			
XX				
PR	23-AUG-2000; 2000US-00649167.			
XX				
PA	(HYSE-) HYSEQ INC.			
XX				
P1	Dmanac RT, Liu C, Tang YT;			
XX				
DR	WPI: 2001-619362/73.			
XX	N-PSDB; AAS78187.			
XX				
PT	New isolated polynucleotide and encoded polypeptides, useful in			
PT	diagnostics, forensics, gene mapping, identification of mutations			
PT	responsible for genetic disorders or other traits and to assess			
PT	biodiversity.			
XX				
PS	Claim 20; SEQ ID NO 44359; 103bp; English.			
XX				
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)			
CC	sequences. (I) is useful as hybridisation probes, polymerase chain			
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,			
CC	and in recombinant production of (II). The polynucleotides are also used			
CC	in diagnostics as expressed sequence tags for identifying expressed			
CC	genes. (I) is useful in gene therapy techniques to restore normal			
CC	activity of (II) or to treat disease states involving (II). (II) is			
CC	useful for generating antibodies against it, detecting or quantitating a			
CC	polypeptide in tissue, as molecular weight markers and as a food			
CC	supplement. (II) and its binding partners are useful in medical imaging			
CC	of sites expressing (II). (I) and (II) are useful for treating disorders			
CC	involving aberrant protein expression or biological activity. The			
CC	polypeptide and polynucleotide sequences have applications in			
CC	diagnostics, forensics, gene mapping, identification of mutations			
CC	responsible for genetic disorders or other traits to assess biodiversity			
CC	and to produce other types of data and products dependent on DNA and			
CC	amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic			
CC	amino acid sequences of the invention. Note: The sequence data for this			
CC	parent did not appear in the printed specification, but was obtained in			
CC	electronic format directly from WIPO at			
CC	ftp.wipo.int/pub/published_pct_sequences			
XX				
SQ	Sequence 598 AA;			
Alignment Scores:				
Pred. No.:	0.000699	Length:	598	
Score:	133.00	Matches:	40	
Percent Similarity:	34.0%	Conservative:	8	
Best Local Similarity:	28.4%	Mismatches:	37	
Query Match:	14.4%	Indels:	56	
DB:	4	Gaps:	5	
US-10-071-510A-16 (1-493) x ABG14000 (1-598)				

Qy	411	CCCGTACCAACCTCTTGACTGAGGCGCTGAGCCCTTCTCCAGACCCCTGCTAACAAGAC	352
		:::	
Db	130	ProteinProserProserProserAla1leProserLeuProProPro-----	145
Qy	351	TCACGCGTCGACGATCCTGGGAATAGATCCATGGATTGATGAGAAATCCTGACTTTG	292
		-----	
Db	145	-----	145
Qy	291	GGATGTTGTTGGTGA AAAACCATTTTCTTCCCGAATCCACTTCCATCTCTGCAGATGCC	232
Db	146	-----SerProginProleuProPro-----	152
Qy	231	AGTTTCGATGAGCGCTCAGACACTTCTCTCCAGACGCGCTCAAGCCCGCAGACTCTG	172
		-----ProProProSerSerProProProSerLeu	162
Db	153	-----	
Qy	171	CCAGCTCCCTGAGCTCTCTCTCCAGACA--CGGCGACACCTCCGGAGAAAGACTTCTCA	115
		:::	
Db	163	ProserProleuLeuProProProProleuSerSerSerProSerSerProleuSerPro	182
Qy	114	TCACGACCAACCCCTCGCTTCACCAAGG-----ACAGCTGGGCTT	73
Db	183	SerProProProProserProProProSerLeuProProSerProProProSerProPro	202
Qy	72	CCCTTCGCGGAATTCCTCCAGCGCTCTCAAACTCGGCTCTTGGGACTCGGATCCC	13
Db	203	ProProProProProginProProSerProProSerSerProleuSerSer-----Pro	220
Qy	12	CCG 10	
Db	221	Pro 221	

RESULT 11

ADFe0423 standard; protein; 598 AA.

ADFe0423; 12-FEB-2004 (first entry)

Human contig polypeptide sequence SEQ ID NO:2790.

biological activity; genetic engineering; hybridisation probe; oligomer; primer; chromosome mapping; gene mapping; recombinant protein production; human.

Homo sapiens.

MO2003080795-A2.

02-OCT-2003.

09-AUG-2002; 2002MO-US025485.

09-AUG-2001; 2001US-0311261P.

(HYSE-) HYSEQ INC.

Tang YT, Yang Y, Wang Z, Weng G, Ma Y;

WPI; 2003-876918/81.

N-PSDB; ADF59971.

New polynucleotides, useful as hybridization probes, oligomers or primers, for chromosome or gene mapping, for the recombinant production of proteins, and for generating antisense DNA or RNA.

Example 3; SEQ ID NO 2790; 571bp; English.

The present invention describes isolated polynucleotide sequences (I), which encode polypeptides (II) with biological activity. Also described: (1) a vector comprising (I); (2) an expression vector comprising (I); (3)

CC The present invention describes isolated polynucleotide sequences (I),  
CC which encode polypeptides (II) with biological activity. Also described:  
CC (1) a vector comprising (I); (2) an expression vector comprising (I); (3)

a host cell genetically engineered to comprise (1) which is operatively associated with a regulatory sequence that modulates expression of (1) in the host cell; (4) a polypeptide (11) encoded by (1); (5) a composition comprising the polypeptide of (4) and a carrier; (6) an antibody directed against the polypeptide of (4); (7) detecting (1) or the polypeptide of (4) in a sample; (8) identifying a compound that binds to the polypeptide of (4); (9) producing the polypeptide of (4); and (10) a collection of polynucleotides comprising at least one of the polynucleotide sequences (1). The polynucleotides (1) can be used as hybridisation probes, oligomers or primers, for chromosome or gene mapping, for the recombinant production of proteins, and for generating antisense DNA or RNA. The present sequence represents a human contig polypeptide sequence, which is used in an example from the present invention.

**SQ Sequence 598 AA;**

Fragment: Score:	
Pred. No.:	0.000659
Score:	133.00
Percent Similarity:	34.0%
Best Similarity:	28.4%
Best Local Similarity:	14.4%
Query Match:	7
DB:	5
Length:	598
Matches:	40
Conservative:	8
Mismatches:	37
Indels:	56
Gaps:	5

US-10-071-510A-16 (1-493) X ADF60423 (1-598)

Oy	411	CCGCCTACCACTCTTGA	CCTGGGCGTCGGCCCTT	CTCCACAGCCTGTAGACAG	352
		:::			
Db	130	ProteinProSerProSerProSerProl	eIaalleProSerLeuProPro	-----	145
Oy	351	TCACGCGTCGATGCTCGGAA	TAGATCCATGGATTGATGAAAA	TCTGACTTTG	292
Db	145	-----	-----	-----	145
Oy	291	GGATGTTGTGGTGAAA	ACCATTCTTCTTCCC	GAAATCCACTTTCATCTTCGAGATGCC	232
Db	146	-----	SerProGl1Proleu	ProPro-----	152
Oy	231	AGTTTCTGATGAGGCTC	ACAGACTTTCTTCCAGAGAC	CCCGAAGGCCCGCACGACTGTG	172
Db	153	-----	ProProProSerSerProProProSerLeu	162	
Oy	171	CCAGCTCCCTGAGCTC	CTCTCTGCACCA---CGGACAGAC	CCCTCCGGAGAAGACTTTC	CA 115
		:::			
Db	163	ProSerProleu	ProProProProleuSerSerSerProleuSerPro	182	
Oy	114	TCACGACGACCCCTGCGCTT	CACCAAGS-----ACAGCTGGG	CCT 73	
			:::	:::	
Db	183	SerProProProProSerProProProSerLeu	ProProProSerProProProSerProPro	202	
Oy	72	CCTTCTCCGGGAATTC	GCCACAGCCTCAAACTCGGCTT	TGGAGATCGGACATCC	13
			:::	:::	
Db	203	ProProProProProGl1ProProSerProProSerSerProleuSerPro	-----Pro 220		

RESULT 12  
ADX68858  
ID ADX68858 standard; protein; 760 AA

KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content.

OS	Unidentified.
XX	
PN	US2004034888-A1.

PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX  
DR WPI; 2004-180133/17.

CC The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US Patent Office at  
CC <http://seqdata.uspto.gov/sequence.html?DocId:2004034888>. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactamannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
CC invention.

sq	Sequence	760 AA
Alignment Scores:		
pred. No.:	0.000849	length: 760
Score:	132.50	Matches: 53
Percent Similarity:	37.5%	Conservative: 7
Best Local Similarity:	33.1%	Mismatches: 43
Query Match:	14.4%	Indels: 57
DB:	8	Gaps: 8

US-10-071-510A-16 (1-493) X ADX68858 (1-760)

[illegible]

```

Db 96 -----ProP SerSerThrProProThrLeuProProSerProProasp 112
Qy 234 GCCAGTTTCTGATGAGCTCAGCAGACTTTCTTCAGACGCTCA-----AGCCCGCC 181
Db 113 -----SerProProLeuProProAlaSerProThrProProVal 126
Qy 180 AGGACTCTGCCA-----GCTCCCTGAGCTCTCTCTGACCACCG 142
Db 127 ThrThrSerProSerProProValThrThrSerProSerProSerProAlaThrThr 146
Qy 141 CAGCACCTCCGAGAGACTTCTCCATCA-----CCAGCCAGCCT 100
Db 147 SerProProProAlaGlnThrProProSerLeuProAsnLeuSerProProSerProPro 166
Qy 99 GCG-----CTTCCAGCAGGAGACAGCTGGGCTCTCT-----70
Db 167 AlaGlySerProProProGlnSerProProThrThrLeProProProSerArgProLe 186
Qy 69 TCTCCGGGAATTCTGCCAGCAGCTCTCAACTCGGCTCTTGGGACTCGGCATCCCCG 10
Db 187 SerProSerProProProAlaAsnProArgProProThrThrArgThrProPro 206

```

## RESULT 13

AAW31852  
ID AAW31852 standard; protein; 763 AA.

AAW31852;

25-MAR-2003 (revised)  
27-APR-1998 (first entry)

Mycobacterium tuberculosis 74 kDa protein.

Tuberculosis; mycobacteria; infection; diagnosis; antimycobacterial;  
antibiotic; vaccine.

Myobacterium tuberculosis.

W09741252-A2.

06-NOV-1997.

18-APR-1997; 97MO-EP001973.

29-APR-1996; 96DE-01017184.

(GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.

Singh M, Honisch C, Esplicita C, Moreno C;

WPI; 1997-549750/50.

N-PSDB; AAT93610.

New DNA and related proteins or RNA derived from M. tuberculosis - used  
for diagnosis of mycobacterial infections, monitoring vaccination and  
development of anti-mycobacterial agents.

Claim 5; Fig 13; 55pp; English.

This novel 74 kDa protein is encoded by an open reading frame of a  
Mycobacterium tuberculosis DNA fragment (see AAT93610) containing  
polymorphic GC-rich sequences. Its amino acid sequence shows a high  
proline content, but there is no homology to any known proline-rich  
antigens of mycobacteria. Novel M. tuberculosis proteins (see AAW31851-  
57) are claimed. These can be produced as recombinant proteins,  
especially in bacterial, yeast, fungal or higher eukaryote host cells,  
and used for diagnosing tuberculosis and other mycobacterial infections  
in humans or animals. The claimed proteins can also be used for  
epidemiological studies, for monitoring vaccination, and for the  
development of vaccines and anti-mycobacterial drugs. (Updated on 25-MAR-  
2003 to correct PR field.)

sq Sequence 763 AA;

Alignment Scores:

Pred. No.:	0.00107	Length:	763
Score:	131.50	Matches:	47
Percent Similarity:	32.9%	Conservative:	10
Best Local Similarity:	27.2%	Mismatches:	53
Query Match:	14.2%	Indels:	64
DB:	2	Gaps:	7

US-10-071-510A-16 (1-493) x AAW31852 (1-763)

```

Qy 410 CCTACCCAGCTCTTGACCTGAGGCT-----GGCCCTCTCCACAGC 366
Db 151 ProPheProProAlaAlaLeuAsnProProAlaProProAlaProProLeuAlaAsn 170
Qy 365 -----CCTGCTAGACAGACTCAGCGCTGACGATGCTGGGAATAGATCAT 318
Db 171 ProProLeuProProAlaProProThrProAlaGlyThrProProAlaAlaPro-----188
Qy 317 GGGATTGATGAGAAATCCTGACTTTGGGATGTTGTGTGA-----AAAGCA 271
Db 189 -----TTPProProValProAlaAlaProLysSerLysPro 200
Qy 270 TTTTCTCCCGAATCCACTTCATCCTGTCAGATGCCAGTTTCTGATGAGCTCAGCA 211
Db 201 AlaSerProProArgPro-----206
Qy 210 GACTTTCTTCAGACAGCTCAAGCGCCGACAGACTCTGCCAGCTCCCTGAGCTCTCT 151
Db 207 -----ProAlaProPheProMetGluPheProProLeuProPro 223
Qy 150 GCACCGAGCGACAGACCCCTCCGAGAGACTTCTCATGACGACGACGCTGGCTTCA 91
Db 224 ValProProAspProLysSerLysGluThrProProAlaProProAlaProProLysPro 243
Qy 90 CCAGGAGACAGCTGGG-----CCTCTCTCTCCGGAATTCAGCA 52
Db 244 ProAlaProProAlaProLysProProValProProLeuProProValProAlaLysLysPro 263
Qy 51 CCAGCTCTCAAACT-----37
Db 264 ProAlaProProAlaProProValAlaValAlaValAlaValAlaProCysArgProPro 283
Qy 36 ---CGCCCTCTGGGACTCGGATCCCGGGCGCGCG 1
Db 284 LeuProProLeuProAsnAsnHisProProAlaProPro 296

```

## RESULT 14

ABB91504  
ID ABB91504 standard; protein; 708 AA.

ABB91504;

31-MAY-2002 (first entry)

Herbicidally active polypeptide SEQ ID NO 715.

Herbicidal; plant; agriculture; herbicide.

Arabidopsis thaliana.

W0200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP009892.

28-AUG-2001; 2001WO-EP009892.

(FARB ) BAYER AG.

Tietjen K, Weidner M;







Qy 57 CTGCCACGACCTCTCAAACTCGGCTTTGGGACTCGGCATCCCCGGGCGCG 1  
|||  
::: |||  
Db 99 SerProProProAlaThrProProProValAlaThr-----ProProProAlaPro 115

Search completed: April 29, 2006, 03:41:52  
Job time : 142 secs

**THIS PAGE LEFT BLANK**

340 CGATGCCCTGGATAGGATCCATGGGATTGATGAGAAATCC

Db	41	-----	41
Qy	280	GTGAACCATTTTTCCTCCCGAATCCACTTCATCCTCTGCGAGATGCCAGTTTCTGATG	221
Db	42	----ProProLeuSerSerProProProLeuProSerProProProLeuSerAlaProTh	60
Qy	220	AGGCTCACACAGACTTCTTCAGCAGCGCTCAAGAGCGCCGACAGACTGTCGACGCTCCC-	163
Db	60	rAlaSerProProProLeuProProValGluSerProProProIleGluSerProPr	80
Qy	162	-----TGAGCTCTCTGCACACAGCGACGACCTT-----	133
Db	80	oProProLeuLeuGluSerProProProProProProLeuGluSerProSerProProSerPr	100
Qy	132	-----CCGAGAAAGACTTCTTC	116
Db	100	oHieValSerAlaProSerGlySerProProProLeuProPhleuProAlaIysProSerPr	120
Qy	115	ATCACGACGACGCTT-----GGGCTTCACACAGGAGACAGTGGGCTCTTCTTC	65
Db	120	oProProSerSerProProSerGluThValProProGlySerThrIleSerProProPr	140
Qy	64	GGGAATTTGGCAACGAGCTCAAACTCGGCTCTTGGAGCTGGCATCCCGC	10
Db	140	oArgSerLeuProSerGluSer---ThrProProValSerThrAlaSerProPro	157

```

Oy      42  AAACCTGAG-----CCTCTGGGACATCCCGGACCGGCG 1
Db      415  ArgSerArgAlaGlyGlyProPheLeuGlyThrArgProProProProPro 432

RESULT 3
T29293
hypochemical protein C50F7.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29293
R:Johnson, D.; Stellyes, L.
Submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C50F7.
A:Reference number: Z20601
A:Accession: T29293
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-309 <J0H>
A:Cross-references: UNIPROT:Q18751; UNIPARC:UPI0000060F74; EMBL:U41557; P1DN:AAA83301.1;
C:Genetics:
A:Gene: CBSP:C50F7.5

Alignment Scores:
Pred. No.:      0.00408      Length:      309
Score:          133.50      Matches:      52
Percent Similarity: 36.8%      Conservative: 8

```

```

RESULT 2
S22697
extension - Volvox carteri (Fragment)
C:Species: Volvox carteri
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
C:Accession: S22697; S21006
R:Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
EMBO J. 11, 2055-2062, 1992
A>Title: A novel extension that may organise extracellular matrix biogenesis in Volvox ca
A:Reference number: S22697; MUID:92289669; PMID:1600938
A:Accession: S22697
A:Molecule type: mRNA
A:Residues: 1-464 <HAL>
A:Cross-references: UNIPROT:Q41645; UNIPARC:UPI000009P7C8; EMBL:X65165; NID:g21991; PIDD
C:Keywords: glycoprotein

Alignment Scores:
Pred. No.:          0 00348      Length:         464
Score:              134.50       Matches:        40
Percent Similarity: 40.7%        Conservative:   40
Best Local Similarity: 33.9%     Mismatches:    33
Query Match:        14.6%        Indels:         37
DB:                 2           Gaps:           5
US-10-071-510A-16 (1-493) x S22697 (1-464)

OY 273 CCAATTTTCCCGGAATTCACCTTCCTGTCAGATGCCAGTTTGATGAGGCTCA 214
    ||| |||||:::||| ||| |||
Db 325 ProArgSerSerProSerProProProProProProPro----- 339
    213 GCAGACTTCTTCCGACGACCTCAAGCCGCGCACGACTTCGCAGTCCGTGAGCT--- 157
    ||| |||||:::||| |||||::: ||| ||| ::|
Db 340 -----SerProProProProArgProSerProSerProProProProArgSerSer 356
    156 -----CCTCTGCACACGGCGACGACCCTTCG 130
    ||| |||||:::||| ||||| ||||| |||
Db 357 ProSerProProProProValValSerProProProProProProArgAlaSerProPro 376
    129 GAGAAGATTCTCCATCAACGACGACCGCTGGCTTCCACACGAGGACGAGGCGTCTCT 70
    :||| |:::||| ||| ||| |||||::: ||| |||
Db 377 ProProProAlaSerSerProProProProProProProProProProProProPro 396
    69 TCTCCGGAAATTCGCGACAGCC-----TCT 43
    ||||| |||||:::||| ||| |||
Db 397 SerProPro-----ProProAlaThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 414

```

Query Match:	14.5%	Indels:	42
D8:	2	Gaps:	7
US-10-071-510A-16 (1-493) x T29293 (1-309)			
QY	467 CCAAGTCTCAGACGACGATGTGGCCCAAGACGAGGAGGGCGCTTGCTGCAGACCCC	408	
	::     ::     ::		
D8	151 ProdlProValAspProSerGIuaSpProGlnProSerValGIuProSerGIuaSpPro	170	
QY	407 TACCACCTCTTGACCTGGGGCCCTGGGCCCTTCGCCACAGCCCTGTATACAGACTCAC	348	
D8	171 GlnProSerelYProProSerProGIuProValAspProSerGIuaSpProGlnProSer	190	
QY	347 GCCTGCAGCATGCTCGGGAAATAG-GATCCATGGGATTGATGAGAAATCTGACTTTGGGA	289	
D8	191 GIySerSerSerProGIyProValAspProSerAsp-----	202	
QY	288 TGTTGTTGGTGAACAACCATTTTCTCCCGAATCCATTCCATCTCTGCAGATGCCAGT	229	
	:		
D8	203 -----GluProSerProSerGIyProProSerProGIyProValAspProSer	218	
QY	228 TTGTGATGAGGCTCAGCAGACTTCTTCAGACAGCTCAAAGCCCCGACGACTTCCCA	169	
D8	219 -----GluAspPro	221	
QY	168 GCTCCCTGAGACTCCTCTGTGACCAACGCGACACACCCTCGGAGAATACTTCCATACCA	109	
	::     ::     ::		
D8	222 LysProSerelYProProSerProGIyProValAspPro--SerAspGIuProSerPro	240	
QY	108 GCCAGCCCTCGCTTCACACGAGGACAGCTGGGCTCTCT-----CCGGGAATT	58	
	::		
D8	241 SerAspProProGIyProProGIyProProGIyProProThrArgArGProProGIy---	259	
QY	57 CTGGCACACGCTCTCAACTCG-----CCTTTGGG-----ACTCGGCAT	16	
D8	260 ---ProProGIyProProThrArgArGProProGIyProProGIyProProThrArgArG	278	
QY	15 CCCCCGGGC 7		
D8	279 ProProGIy	281	

RESULT 4  
D96711  
hypothetical protein F24J5.8 [imported] - Arabidopsis thaliana  
[Species: Arabidopsis thaliana (mouse-ear cress)]  
[Date: 02-Mar-2001 #sequence revision 02-Mar-2001 #text change 09-Jul-2004]

C:\Accession: D96711  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, U.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L., Jenkins, J., Johnson-Hopson, C., Khan, S., Khaykin, E., Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Matli, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.; Unayam, L.; Tallon,  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: D96711  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-708 <STO>  
A:Cross-references: UNIPROT:Q9SXX1; UNIPARC:UPI0000048426; GB:AE005173; NID:G5734709; PD

C:Genetics:  
A:Gene: F2J45\_8  
A:Map position: 1

Alignment Scores:

Alignm. No.:	Score:	Length:	Matches:
Percent Similarity:	130.50	708	37
Best Local Similarity:	41.1%	Conservative:	9
Query Match:	33.0%	Mismatches:	39
	14.1%	Indels:	27
		Gaps:	5

DB: 2

US-10-071-510A-16 (1-493) x D96711 (1-708)

OY 273 CCATTTCTTCCTCCCG-----AATCCATTTCATCTCTGCAGATGCCAGTTTCTGA 223  
|||:::||  
Db 29 ProAlaInProProProValThnSerProLeuProProSerAlaPro----- 44  
OY 222 TGAGGCTCAGACACTTCTTCGACGAGCCTCAAGCCCCGGCAGACTCTGCAGTCCC 163  
:::||||  
Db 45 -----ProProAlaInProProProProProProProProValThnThnSerProPro 62  
162 -----TGAGCTCTCTGTGCACCACGGAGCACCTCCGGAAGAATTCT----- 118  
|||||  
Db 63 ValAlaInProProProProProProProProProProProProProProProPro 82  
117 CCATCACCAGCAGCAGCCCTGCGCTTCCACAGGAGACAGCTGGGCTCTCTCT----- 67  
|||:::||  
Db 83 ProGlnProValIleProSerProProSerProProSerThnSerProProGlnProValIle 102  
OY 66 -----CGGGAATTCGCGACGAGCAGCTCTCAACT 37  
|||:::||  
Db 103 ProSerProProProSerAlaSerProProProProAlaInProProLeuProSerSer 122  
OY 36 CGGCTCTTGGAGACTCGGCATCCCCCGGCGCCG 1  
|||  
Db 123 ProProProProAlaSerValProProProProArgPro 134

RESULT 5  
F84522  
Probable proline-rich protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: F84522  
R:Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Unayam, L.; Tallon, L.  
aus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
ature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420; MUID:20083487; PMID:10617197  
A:Accession: F84522  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-191 <STO>  
A:Cross-references: UNIPROT:O82327; UNIPARC:UPI000009PF0P1; GB:AE002093; NID:g3650031; PD

C:Genetics:

A:Gene: At2g14890  
 A:Map position: 2  
 C:Superfamily: proline-rich protein 3

**Alignment Scores:**  
 Pred. No.: 0.00763 Length: 191  
 Score: 130.00 Matches: 44  
 Percent Similarity: 36.7% Conservative: 7  
 Best Local Similarity: 31.7% Mismatches: 38  
 Query Match: 14.1% Indels: 50  
 DB: 2 Gaps: 7

US-10-071-510A-16 (1-493) x F84522 (1-191)

QY	405	CCCACTCTTACCTGGGGCTGGCCCTTCTCCACAGCCCTGTAGACAGACTCAGC	346
Db	23	ProthrSerProProThrAlaThrProAlaProProThrProthr	37
QY	345	GTCCAGCATGCTGGGAATGAGATCCATGGATGTATGAGAAATCTGACTTTGGGATGT	286
Db	38	-----ThrPro-----	39
QY	285	TGTTGTGTAACCAATTTCTTCCCGCAATCCACTTCATCTGTGAGATGCCAGTTTC	226
		:    :    :	
Db	40	-----ProProAlaAlaThrProProAla-----	48
QY	225	TGATGAGCTCAGCAGACTTTCTTCAGCAGCCTCAAGCCCGCAGACTCTGCCACT	166
		:    :    :	
Db	49	-----SerAlaProProProAlaThrThrThrSerProPro	59
QY	165	CCC-----TGAGCTCTCTCTGCAGCCAGCAGCAGCCTCCGGAGAAAGACTTTCATCA	112
		:    :	
Db	60	ProValThrThrAlaProPro-----ProAlaAsnProProProProAlaSerSerProPro	78
QY	111	CCAGCCAGC-----CTGCGCTTCCACCGAGGACAGCTGGGCTCTTCTCCGGGAATT	58
Db	79	ProAlaSerProProProAlaThrProProProAlaSerProProProProAla	98
QY	57	CTGCCACCGCCTCTCAAACTCGGCTTTGGAGCTGGGATCCCGCGGCGCG	1
		:    :    :    :    :	
Db	99	SerProProProAlaThrProProProAlaAlaThr-----ProProProAlaPro	115

**RESULT 6**  
 T04859  
 extensin homolog F28A21.80 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_rev150n 23-Apr-1999 #text\_change 09-Jul-2004  
 C:Accession: T04859  
 R:Bayan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15387  
 A:Accession: T04859  
 A:Molecule type: DNA  
 A:Residues: 1-839 <BEV>  
 A:Cross-references: UNIPROT:Q9SN46; UNIPARC:UPI00000A17EF; EMBL:AL035526  
 A:Experimental source: cultivar Columbia; BAC clone F28A21

**Alignment Scores:**  
 Pred. No.: 0.0103 Length: 839  
 Score: 129.00 Matches: 49  
 Percent Similarity: 40.8% Conservative: 20  
 Best Local Similarity: 29.0% Mismatches: 54  
 Query Match: 14.0% Indels: 46  
 DB: 2 Gaps: 7

US-10-071-510A-16 (1-493) x T04859 (1-839)

QY	404	CCACTCTTGAACCTGGGG-----CCTGGC-----	381
		:    :    :	

Db 356 ProProValaCyLeuGlyLeuProGlyPheAspAArgArgAsnCysLeuProAlaArg 375  
QY 360 CCCTTCTCCACAGCCCTGCTAGACAGACTCAGCGCTGACAGATGCTGGGAATA----- 326  
Db 376 ProAlaGlnArgSerProGlyGlnCysAlaAlaPheSerSerLeuPro-ProValaAspCy 395  
QY 325 -GGATCCAGGATGATGAGAAATCCGACTTGTGGATGTTGGTGAATAAACATTTT 267  
Db 395 GblySerPheGlyCyBglyArgSerThrArgProProValaValPro-SerProIot 415  
QY 266 CTTCCCGGATCCACTTCATCCTCTGAGATGCCAGTTTCTGATGAGCCAGCAGACT 207  
Db 415 hThrProSerPro-----GlyGlySerPro 424  
QY 206 TTCTTTCACAGCCTCA---AGCCCGGACAGACTCTGCCAGCTCCCTAGCTCTCTCG 150  
Db 424 roSerProSerIleSerProSerProIleThrAlaProSerProThrProIos 444  
QY 149 CACCAAGGAGAGACCCCTCGGAG-----AMGACTTCAT 114  
Db 444 erProGlyGlySerProProSerProSerIleValProSerProProSerThrProIos 464  
QY 113 CACCAAGGAGACCCCTCGGCTTCCACAGGAGAGCTGGCCT----- 73  
Db 464 erProGlySerProProThrProThrProThrProThrProGlySerProProIos 484  
QY 72 -----CCTTCTCCGGGAATTTCTCCACAGCCTCTCAAACTCGGCTTTGGGA 24  
Db 484 erProThrThrProThrProThrProGlySerProProSerSerProThrProThrPro 504  
QY 23 CTCGGGATCCCGGGCGGCGCG 1  
Db 504 lyGlySerProProSerSerPro 511

## RESULT 7

T50568  
probable multi-domain regulatory protein [imported] - Streptomyces coelicolor  
CISpecies: Streptomyces coelicolor  
C.Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #ext\_change 09-Jul-2004  
C.Accession: T50568  
R.Redenbach, M.; Kleier, H.M.; Denapate, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopw  
Mol. Microbiol. 21, 77-96, 1996  
A.Title: A set of ordered contigs and a detailed genetic and physical map for the 8 Mb S  
A.Reference number: Z20556; MUID:97000351; PMID:8843436  
A.Accession: T50568  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-1334 <RED>  
A.Cross-references: UNIPROT:Q9RKR9; UNIPARC:UPI00000DB642; EMBL:AL133220; PIDD:CA61705.  
A.Experimental source: strain A3(2)  
C.GeneticB:  
A>Note: SCC75A.05C

## Alignment Scores:

Pred. No.:	0.0106	Length:	1334
Score:	129.00	Matches:	49
Percent Similarity:	39.0%	Conservative:	13
Best Local Similarity:	30.8%	Mismatches:	59
Query Match:	14.0%	Indels:	38
DB:	2	Gaps:	7

US-10-071-510A-16 (1-493) x T50568 (1-1334)

QY 462 TCCTCAGACAGACATGTTGCTCCACAGAGAGGCGCTGTGCT-----TGAGACCC 409  
Db 300 AlaSerAlaProAlaValAlaProGlySerGlyGlyProAlaProGlyTPrPro 319  
QY 408 CTAACCACTCTTGACCTGGGCGTGGCCCTTCTCCACAGCCCTGCTAGACAGCTCA 349  
Db 330 AlaProGlyThrAlaProGlySerSerThrAlaProProIleAsp-----ThrAlaSer 337  
QY 348 CCGCTCAGACATGCTCGGGAATAGATCCATGATGATGAGAAATCTGACTTTGGGA 289

Db 338 AlaAlaSerThrAla----- 342  
QY 288 TGTGTGTGTGAAAAACATTTTCTTCCCGGAATCCACTTCCATCTCTGAGATGCCAGT 229  
Db 343 -----ProAlaProGlyProThrSerAlaProGlyThrAlaProAla 357  
QY 228 TTCTGATGAGCTCAGACACTTTCTTCAGCAGGCTCAAGCCCGGACAGACTCTGCCA 169  
Db 358 -----GlyThr-----AlaAlaProAlaProGlyThrAlaGlyPro 369  
QY 168 GCTCCCTGAGCTCTCTGACACAGCAGGAGACCTCCGGAGAAAGACTTCTCATCACA 109  
Db 370 AlaProGlyThrSerThrAlaProGlyThrAlaProValAlaGlyThrProAlaPro 389  
QY 108 GCCAGCCCTGGCTTCCACAGGAGACAGTGGCTCTCT-----TTCGCG 64  
Db 390 GlyThr-----AlaProAlaProGlyThrAlaGlyProAlaArgAspThrSerThrAlaPro 408  
QY 63 GGAATTCGCCACAGCCTCTCAAACTCGGCTCTTGGAGCTCGGATCCCGGCG 7  
Db 409 GlyThrAlaProValAlaGlyThrThrProAlaProGlyThrAlaProAlaProGly 427

## RESULT 8

S57447  
HPBRII-7 protein - human  
N.Alternate names: HPBRII-4 protein  
CISpecies: Homo sapiens (man)  
C.Date: 10-Oct-1995 #sequence\_revision 03-Nov-1995 #ext\_change 05-Oct-2004  
C.Accession: S57447; S57489  
R.Fleischhauer, K.L.  
submitted to the EMBL Data Library, June 1992  
A.Reference number: S57447

A.Accession: S57447  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-551 <FLR>  
A.Cross-references: UNIPROT:Q16630; UNIPARC:UPI000006D566; EMBL:X67336; NID:g871300; PIDD  
A.Accession: S57489  
A.Status: preliminary  
A.Molecule type: mRNA  
A.Residues: 1-551 <FLR>  
A.Cross-references: UNIPARC:UPI000006D566; EMBL:X67337; NID:g871298; PIDD:CAA47752.1; PIT  
C.GeneticB:  
A.Introns: 231/3  
F:82-151/Domain: ribonucleoprotein repeat homology <RRM4>

## Alignment Scores:

Pred. No.:	0.0145	Length:	551
Score:	127.00	Matches:	67
Percent Similarity:	35.9%	Conservative:	7
Best Local Similarity:	32.5%	Mismatches:	77
Query Match:	13.8%	Indels:	56
DB:	2	Gaps:	10

US-10-071-510A-16 (1-493) x S57447 (1-551)

QY 467 CCAATCTCAGACAGACATGTTGCCCCAGACAGGA----- 429  
Db 189 ProGlyGlySerSerArgAlaAlaPheProGlnGlyArgGlyArgPhePro 208  
QY 428 GGGGCTGTGCTGAGACCCCTACCCACTCTTGAACCTGGGCGCTGAG----- 381  
Db 209 GlyAlaValaProGlyGlyAspArgPheProGlyProAlaGlyProGlyGlyProProPro 228  
QY 380 CCGTTCTCCACAGCCTGCTAGACAGACTCAGCGCTGAGAGTCCCTGGGA---ATAGG 324  
Db 229 ProPhe-----ProAlaGlyGlnThr-ProProArgProProLeuGlyProProGly 245  
QY 323 ATCATGAGATGATGAGAAATCCGACTTGTGGATGTTGTGGAAACCATTTCTT 264  
Db 245 yProProGlyProProGlyProProGlyGlnValleuPro-ProProleuAlaG 265  
QY 263 CCGCG-----AATCACTTCATCTCTGACAGATGCCAGTTTCTGATGAG 219



Db 265 l yProProba nA rG l yA sP a rP r oP r o - P r oV a i l l e uP h eP r oG l y i nP r oP h e 284

Qy 218 G C T C A G A C A C T T T C C A G C A G C C T C A A G C C C G C C A G - A C T C T G C A G C T - - - - - 166

Db 285 G l y G l n P r o P r o l e u G l y P r o l e u P r o P r o G l y P r o P r o P r o P r o P r o P r o l a P r o l y T y r G l y 304

Qy 165 - - - C C C T G A G C T C C T C C T G C A C C A G G C A G C A C C C T C C G A G A A G A C T T C T C C A T C A C C A 109

Db 305 P r o P r o P r o G l y P r o P r o P r o P r o G l n G l n l y P r o P r o P r o P r o P r o P r o G l y P r o P r o P r o 324

Qy 108 G C C A G C C C T - - - - - G G C C T T C C A C C A G C A G C A G C T 79

Db 325 P r o A r g P r o P r o G l y P r o l e u G l y P r o P r o l e u T h r l e u A l a P r o P r o P r o H i s l e u P r o 344

Qy 78 G G G C C C C T T C T C C G G A A T T C T G C A C A G C C T C T C A A C T C G G C C T C T T - - - - - 28

Db 345 G l y P r o P r o - - - P r o G l y A l a P r o P r o P r o l a P r o H i s V a l A s n P r o l a P h e P r o 363

Qy 27 - - - - - G G A C T C G G C A T C C C C G - - - - - 10

Db 364 P r o P r o T h r A s n S e r G l y M e t P r o T h r S e a S p e r A r G l y P r o P r o P r o T h r A s P r o 383

Qy 9 - - - G G C C G G C C G 1

Db 384 T y r G l y A r g P r o 387

RESULT 9

A47655

spliceosome-associated protein SAP 62 - human

C:Species: Homo sapiens (man)

C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 09-Jul-2004

C:Accession: A47655

R:Bennett, M.; Reed, R.

A:Title: Correspondence between a mammalian spliceosome component and an essential yeast

A:Reference number: A47655; MUID:94023929; PMID:8211113

A:Status: preliminary

A:Accession: A47655

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-464 <BEN>

A:Cross-references: UNIPROT:Q15428; UNIPARC:UP1000016AF5; GB:I21990; NID:g409218; PIDN:

C:Gene: GDB:SAP62

A:Cross-references: GDB:1381638; OMIM:600796

A:Map position: 19p13.3-19p13.3

C:Keywords: tandem repeat; zinc finger

Alignment Scores:

Pred. No.: 0.0158 Length: 464

Score: 126.50 Matches: 52

Percent Similarity: 38.8% Conservative: 12

Best Local Similarity: 31.5% Mismatches: 66

Query Match: 13.7% Indels: 35

DB: 2 Gaps: 9

US-10-071-510A-16 (1-493) x A47655 (1-464)

440 C C C A G A C A G G G A G G G C C T G T G C T G A G A C C C C T A C C A C C T T T G A C C T G G G G C C T G G C 381

Qy 440 C C C A G A C A G G G A G G G C C T G T G C T G A G A C C C C T A C C A C C T T T G A C C T G G G G C C T G G C 381

Db 265 P r o P r o T h r G l y P r o l a P r o S e r G l y P r o P r o G l y P r o P r o G l n l e u P r o P r o P r o l a 284

Qy 380 C C C T T C C C A C C C C T G T A G A C A G A C T C A C G G C T G C A G C A T G C C T G G A A T A G A T C 321

Db 285 P r o G l y V a l H i s P r o P r o l a P r o V a l V a l H i s P r o - - - P r o l a S e r G l y V a l 301

Qy 320 C A T - - - - - G G A T T G A T A G A A A C C T G A C T T T G G A T G T T G T G T G A A A C 273

Db 302 H i s P r o P r o l a P r o G l y V a l H i s P r o P r o l a - - - - - P r o G l y V a l H i s 316

Qy 272 C A T T T T C T C C C C A A T C A C - - - - - T T C A C T C T G C A G A T G C C A G T 228

Db 317 P r o P r o l a P r o G l y V a l H i s P r o P r o T h r S e r G l y V a l H i s P r o P r o l a P r o G l y V a l 336

QY	227	TCTGATGAGGCTCAGACACTTTCTTCCAGAGGCTCAAGGCGGACAGACTCTGCAG	168
Db	337	HisProProAlaProGlyValHis--ProProAla-----ProGlyValHisProProA	354
QY	167	CTCCCTGAGCT---CTCCTGCACACGAGGACGACCTT-----C	132
Db	354	laProGlyValHisProProAlaProGlyValHisProProProserAlaGlyValHisP	374
QY	131	CGAGAGACTTCTTCATCAACGACGAGCCCTGCGCTTCCACGAGGACAGCTGGG----	76
Db	374	roGlnAlaProGlyValHisProAlaAlaProAlaValHisProGlnAlaProGlyValH	394
QY	75	--CCCTCTTCCGGGAAFTTCGCCACAGCGCTCAACTGGGCTTTGGAGCTCGGC	18
Db	394	isProProAlaProGlyValHisProGlnAlaProGlyValHisPro-----G	410
QY	17	ATCCCCCGGAGC	7
Db	410	lnProProGly	413
RESULT 10			
S13383			
hydroxyproline-rich glycoprotein - sorghum			
C.Species: Sorghum bicolor (sorghum)			
C.Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 09-Jul-2004			
C.Accession: S13383; S14449			
R.Raz, R.; Cretin, C.; Puigdomenech, P.; Martinez-Izquierdo, J.A.			
Plant Mol. Biol. 16, 365-367, 1991			
A.Title: The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum vulgare.			
A.Reference number: S13383; PMID:1693107			
A.Accession: S13383			
A.Molecule type: DNA			
A.Residues: 1-283 <RA21>			
A.Cross-references: UNIPROT:P24152; UNIPARC:UPI000017883B; EMBL:X56010			
A.Note: the source is designated as Sorghum vulgare			
R.Raz, R.			
submitted to the EMBL Data Library, October 1990			
A.Reference number: S14449			
A.Accession: S14449			
A.Molecule type: DNA			
A.Residues: 1-238, 'Y', 240, 'QAH', 244-283 <RA22>			
A.Cross-references: UNIPARC:UPI000012A3A9; EMBL:X56010; NID:921626; PIDN:CAA39485.1; PID			
C.Superfamily: hydroxyproline-rich glycoprotein			
C.Keywords: glycoprotein; hydroxyproline			
Alignment Scores:			
Pred. No.: 0.0324 Length: 283			
Score: 122.50 Matches: 43			
Percent Similarity: 33.5% Conservative: 10			
Best Local Similarity: 27.2% Mismatches: 54			
Query Match: 13.3% Indels: 51			
DB: 2 Gaps: 5			
US-10-071-510A-16 (1-493) x S13383 (1-283)			
QY	473	ACGTGGCAAGTCCTCAGACAGCATGTTGCCACAGACGAGGAGGCGCTGCTGGA	414
Db	130	ThrlyProThrProlyseProProAlaThrlyProProthPrProProValTyThrPro	149
QY	413	GACCCCTACCACTCTTGACTGGGGCCCTGGCCCTTCTCCACAGCCCTGCTAGACAG	354
Db	150	SePrProlyseProProValThrlyseProPro-ThrProlyseProThr-Pro-----	165
QY	353	ACTCAGACGCTCGACGATGCTGGGAATAGATCATGATGATGAGAAATCCTGACTT	294
Db	165	-----	165
QY	293	TGGGATGTTGGTGATAAAACATTTCTTCCCGGAATCATTCTTCTGCAAGATG	234
Db	166	-----ProValTyThrProAlaPro-----	172
QY	233	CCAGTTTTCGATGAGGCTCAGACACTTTCTTCCAGAGGCTCAAGGCCGCGACGACTC	174

```

Db 173 -----LybProProValThirlyProProThHiT 183
Qy 173 TCCAGCTCTCTGAGCTCTCTGACACAGGAGACCTCCGGAAGACT--TCTC 117
Db 183 hPrProSerPro--LysProProThSerLysProProValTyThrProSerP 202
Qy 116 CATCACACACAGCCCTGCGCTTCCACAGGAGACCTGGGCTCTCTCCGGGAATTC 57
Db 202 rOlYProProLysProSerProProThTyThrProThProLysProProAlaThrl 222

Qy 56 TCCACACAGCTCT--CAAACTCGGCTCTTGGAGCTCGGCATCCCGG 10
Db 222 yErProProThSerThrProThHiHPrProLysProThrProHiThrPro 238

RESULT 11
142731
atrophin-1 related protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42731
R:Khan, F.A.; Margolis, R.L.; Loev, S.L.; Sharp, A.H.; Li, S.H.; Ross, C.A.
submitted to the EMBL Data Library, December 1995
A:Description: CDNA sequence and expression of an atrophin-1 (DRPIA disease gene) relate
A:Reference number: 223250
A:Accession: T42731
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1006 <KHA>
A:Cross-references: UNIPROT:Q62901; UNIPARC:UPI0000086F1; EMBL:U44091; NID:g1297310; PI
C:Genetics:
A:Gene: ARP

Alignment Scores:
Pred. No.: 0.043 Length: 1006
Score: 121.50 Matches: 57
Percent Similarity: 31.5% Conservative: 5
Best Local Similarity: 28.9% Mismatches: 50
Query Match: 13.2% Indels: 86
DB: 2 Gaps: 9

US-10-071-510A-16 (1-493) x T42731 (1-1006)
Qy 467 CCAAGTCTCAGAGCAGAGCATGTTGCCACAGAGGGGCTGTGTCGAGACCCC 408
Db 290 ProGlyProHiseSer-----LeuGlnThrGlyProLeuLeuGlnHisPro 304
Qy 407 TACCCACTCTTGACCTG-----GGGCTGGCCCTTCTCC 372
Db 305 GlyProProGlnProHeGlyLeuThrProGlnSerSerGlnGlyGlnGlyProLeuGly 324
Qy 371 CACAGCCCTGTAAGACAGACTACGGGTGACGAGTCCCTGGAAATGATTCATGGGATT 312
Db 325 ProSerProAlaAlaAlaHisProHisSerThr----- 335
Qy 311 GATGAGAATCTGACTTGGGATGTGTGGTGAAGAACATTTCTCCCGAATCCAC 252
Db 335 ----- 335
Qy 251 TTCATCTCTGAGATGCGAGTTTCTGATGAGGCTCAGACAGATTCTTCCAGACAGCT 192
Db 336 -----1leGlnLeuProAlaSerGlnSerAla-----LeuGlnPro 347
Qy 191 CAAGCCCGCCAGACT-CTGCCAGCTCCCTGAGCTCT----- 154
Db 348 GlnGlnProProAlaGlnGlnProLeuProProAlaProLeuAlaMetProHisIleLys 367
Qy 153 -----CCTGCACGACGG-----CAGCACTCTCCG 130
Db 368 ProProProThrThrProIleProGlnLeuProAlaProGlnAlaHisIleHisProPro 387
Qy 129 GAGAAGACTTTCGATCACCAGCCAGC-----CCTGGGCTT 94

```

```

Db 368 HisLeuSerGlyProSerProPheserMetAlaAlaLeuProProProAlaLeu 407
Qy 93 CCACCA-----GGAGACGCTGGGCTCTCTT-----CCGGGAATT 58
Db 408 LysProLeuSerSerLeuSerThrHisHsProProSerAlaHisProProProLeuGln 427
Qy 57 CTGCCACACAGCTCTCAAACTCGGCTCTTGGAGACTCGGCATCCCGCGGC 7
Db 428 LeuMetProGlnSerGlnProLeuProSerSerProAlaGlnProProGly 444

RESULT 12
H96711
hypothetical protein F14K14.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96711
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Comn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiz, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96711
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-222 <STO>
A:Cross-references: UNIPROT:Q95740; UNIPARC:UPI000009C950; GB:AE005173; NID:g6524187; PI
C:Genetics:
A:Gene: F14K14.17
A:Map position: 1

Alignment Scores:
Pred. No.: 0.0423 Length: 222
Score: 121.00 Matches: 42
Percent Similarity: 33.6% Conservative: 8
Best Local Similarity: 28.2% Mismatches: 49
Query Match: 13.1% Indels: 50
DB: 2 Gaps: 5

US-10-071-510A-16 (1-493) x H96711 (1-222)
Qy 411 CCCCTAACCCACTTGTGACTGGGGCTGGCCCTTCTCCACAGCCCTGTAGACAGAC 352
Db 60 ProValSerAlaAlaGlnProProAlaSerProValThrProProProAlaValThrPro 79
Qy 351 TCA-----CGGTGACAGATGCTCGGAAATGATCCATGGGATTGATGA 307
Db 80 ThrSerProProAlaProLysVal----- 87
Qy 306 GAAATCTGACTTGGGATGTGTGTGTAAGAACATTTCTTCCCGGATTCACATTCGA 247
Db 88 -----AlaProValIleSerProAlaThrProPro 97
Qy 246 TCTCTGAGATGCCAGTTTGTGATGAGGCTCAGACAGATTCTTCCAGACCTCAAG 187
Db 98 ProGlnProProGlnSer-----ProProAlaSerAla 108
Qy 186 CCGCGCAGACTTGGCCAGCTTCCCTGAGCTCTCTGACACAGGAGACACCTCCGAG 127
Db 109 ProThrValSerProProProValSerProProProAlaProThrSerProProThr 128
Qy 126 AAGACTTCTCATACACAGCAGCCGCTTCCGTCACAGGAGACAGACCTGGGCT----- 73
Db 129 ProAlaSerProProProAlaProAlaSerProProProAlaProAlaSerProProPro 148
Qy 72 -----CCTTCTCGGGAATTCGACACAGCCCTCT 43
Db 149 AlaProValSerProProProProValGlnAlaProSerProIleSerLeuProProAlaPro 168

```

42 CAAACTCGGCTCTTGGAGTCGCGCAT 16  
Db 169 AlaProAlaPro-----ThryshHis 175

## RESULT 13

hypothetical protein F23H11.22 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: C96623

R:Theslogist, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

anzen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzman, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MIMD:21016719; PMID:11130712

A:Accession: C96623

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-929 <STO>

A:Cross-references: UNIPROT:Q9XIE0; UNIPARC:UPI00000A7E6A; GB:AE005173; MIMD:95080823; PI

A:Gene: F23H11.22

A:Map position: 1

## Alignment Scores:

Pred. No.:	0.047	Length:	929
Score:	121.00	Matches:	64
Percent Similarity:	34.9%	Conservative:	16
Best Local Similarity:	27.9%	Mismatches:	63
Query Match:	13.1%	Indels:	86
DB:	2	Gaps:	13

US-10-071-510A-16 (1-493) x C96623 (1-929)

473 ACCTGGCCAGATCTCTGAGCAGAGCATGTTGCCAGACAGGAGGCGCTGCTGCA 414  
Db 227 SerPheProSerSerProSerGlnIle-----HisSerGlyGlyArgSer--- 242  
413 GACCCCTACCCACTCTTGAACCTGGGCGCTGGCCCTTCTCCACACCCCTGCTAGACAG 354  
Db 243 -----ProProLeuProLeuProProGlyGlnPheThrAlaGlyAaAla---Ser 258  
353 ACTCAGCGCTGACGATGCTCT-----GGGAATAGATCATGCGG 315  
Db 259 PheProSerSerThrGlnProProProGlyGlnTyrMetAlaGlyAaAlaAlaSerPhePr 278  
314 ATTGATGAGAATCTGACTTCTGGATGTTGGTG-----AAAAATTCTTCTCC 261  
Db 278 oSerSerThrProProProProGlyGlnTyrMetAlaGlyAaAlaAlaProPheSerSers 298  
260 CGAATCAGCTTCCA----- 247  
Db 298 eThrProLeuProProGlyGlnTyrProAlaValAaAlaGlnLeuSerThrSerAlaP 318  
246 -----TCCTCTGAGATGCGAGTTCATGATGAGGCT 216  
Db 318 roSerValProLeuProProGlyGlnTyrThrAlaValAaAlaAlaProPheSerThrSert 338  
215 CA-----GCAGACTTCTTCCA-----GCAGCT 192  
Db 338 hGlnProValSerLeuProProGlyGlnTyrMetProGlyAaAlaAlaLeuSerAlas 358  
191 CAAGCCCGCC-----ACGACTCTGCGCAGCTCCCTGAGCTCTCTCT----- 151  
Db 358 eThrProLeuThrProGlyGlnPheThrThrAlaAaAlaAlaProProAlaProProGlyP 378

150 -----GCACCAGCGCAGC 138  
Db 378 roAlaAaGlnThrSerProProProProProProSerAlaAlaAlaAlaProProPro 398  
137 ACCCTCCGAGAGAACTTCTCCATCAGCAGCCCTGCGCTTCACACGAGGAGCA----- 82  
Db 398 roProProProProProProProProProProProProProProProProProProPro 418  
81 -GCTGGGCTCTCTCTCCGGAATTCGCCACGACCTCTCAAACTCGGCTCTTGGGA 24  
Db 418 lYAlaGlyProProProPro-----ProPrometSerLysLys-----GlyP 432  
23 CTCGGCATCCCGCGGCGCGCG 1  
Db 432 roProLysProProGlyAaPro 439

## RESULT 14

probable transcription factor HUA2 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004

C:Accession: T51947

R:Chen, X.; Meyerowitz, E.M.

Mol. Cell 3, 349-360, 1999

A:Title: HUA1 and HUA2 are two members of the floral homeotic AGAMOUS pathway.

A:Reference number: Z25882; MIMD:9214850; PMID:10198637

A:Accession: T51947

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1392 <CH>

A:Cross-references: UNIPROT:Q9XER9; UNIPARC:UPI00000A0734; EMBL:AF116556; PIDDN:AAD31171.1

A:Experimental source: cultivar Columbia

C:Genetics:

A:Gene: HUA2

A:Map position: 5

## Alignment Scores:

Pred. No.:	0.0532	Length:	1392
Score:	120.50	Matches:	51
Percent Similarity:	38.5%	Conservative:	14
Best Local Similarity:	30.2%	Mismatches:	63
Query Match:	13.1%	Indels:	42
DB:	2	Gaps:	10

US-10-071-510A-16 (1-493) x T51947 (1-1392)

386 CCGTGGCCCTCTC-----TCCACAGAC-----CCTGCT 360  
Db 955 ProGlyPhePheSerSerSerThraenPheGluAaPAspGluAaPAspAaPAspLeuProThr 974  
359 AGACAGACTCAGCGCTGACGATGCTGGGAATAGATCCAGATGATGAGAAATCC 300  
Db 975 SerGlnLysGlnLysSerThrSerAlaGlyGluAaValSerAlaLeuAaPAspLeuGlu 994  
299 TGACTTGGGATGTTGGTGAAC-----CATTTCTTCCCGGATTCACCTTCATCCT 243  
Db 995 ---IleHisAspThrSerSerSerLysCyseHisArgValLeuGluAaPAspValaPAspHisGlu 1013  
242 CTGCAGATG---CGAGTTCT-----GATGAGGCTCAGACGACTTCTT----- 202  
Db 1014 LeuGluMetGluAaPAspValSerGlyGlnArgLysAaPAspValAlaProSerSerPheCyseGlu 1033  
201 -----CCAGACGCTTCAGAGCCCGCCAGC 178  
Db 1034 AsnLysThrLysGlnLysSerLeuAaPAspValMetGluProValAlaGluLysSerThrGlu 1053  
177 ACTCTGCGCACTCC---TGAGCTCTCTCTGACACAGCAGACACCTCCGAG----- 127  
Db 1054 PheAsnProLeuProGluAaPAspSerProProLeuProGlnLysSerProProProLeuPro 1073  
126 -----AAGACTCTCATCAGCAGCAGCCTGCGCTTCACAGGAGCAGCTGGGCT 73  
Db 1074 ProLeuProProSerProProProProProProProProProProProProProProPro 1093



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_r2p model

Run on: April 29, 2006, 03:37:25 ; Search time 33.9 Seconds  
(without alignments)  
2052.068 Million cell updates/sec

Title: US-10-071-510A-16  
Perfect score: 923  
Sequence: 1 cggccgcgcgcggggagtgcc.....cctgcctcatcttgagcctgcg 493

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 216643 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ r2p model -DRV=xlp  
-Q=/abs/ABSSWEB\_epool/US10071510/runat\_28042006\_170847\_18707/app\_query.fasta\_1  
-DB=UniProt -QFMT=blast -SUFFIX=rup -MIMATCH=0.1 -LOOFCU=0 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdt -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs802p  
-USER=US10071510 @CCN 1.1 466 @runat\_28042006\_170847\_18707 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOGS -REV TIMOUT=120  
-WARN TIMOUT=30 -THREADS=1 -GAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: UniProt 05.80.\*  
1: uniProt\_sprot.\*  
2: uniProt\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	592	64.1	489	2	Q8N7G8_HUMAN
2	586	63.5	904	2	Q6ZM23_HUMAN
3	404	43.8	895	2	Q8C117_MOUSE
4	404	43.8	975	2	Q4FZC9_MOUSE
5	305	33.0	262	2	Q8BMM1_MOUSE
6	149	16.1	555	1	GPI_CHLRE
7	144.5	15.7	911	2	Q4SM77_TETNG
8	144.5	15.7	4027	2	Q512R0_CHLIN
9	141	15.3	1143	2	Q948Y6_VOLCA
10	140.5	15.2	3889	2	Q6SS88_CHLRE
11	139.5	15.1	3409	2	Q6SS86_CHLRE
12	139	15.1	2371	2	Q58N45_CHLRE
13	138	15.0	636	2	Q7SFI5_NEUCR
14	138	15.0	676	2	Q95UC9_PIG
15	136	14.8	745	2	Q89X06_BRAUA
16	136	14.7	236	2	Q6QJ26_ARATH

17	136	14.7	342	2	Q6ZD62_ORYSA
18	136	14.7	731	2	Q65530_ARATH
19	135	14.6	597	2	Q5E8F5_9BURK
20	135	14.6	659	2	Q6C708_YARLI
21	134.5	14.6	417	2	Q5ZPC7_9DELT
22	134.5	14.6	464	2	Q41645_VOLCA
23	134.5	14.6	566	2	Q95UD1_PIG
24	134.5	14.6	1701	2	Q61DV2_CABBR
25	134	14.5	511	2	Q95UD0_PIG
26	134	14.5	1495	2	Q5K1Y9_CRYNE
27	134	14.5	1525	2	Q55TE9_CRYNE
28	133.5	14.5	409	2	Q18751_CABEL
29	133.5	14.5	368	2	Q41334_TETNG
30	133.5	14.5	687	2	Q948Y7_VOLCA
31	133.5	14.5	1997	2	Q81RM7_CHLRE
32	133	14.4	400	2	Q51HD8_MAGGR
33	132.5	14.4	698	2	Q9ASK4_ORYSA
34	132	14.3	303	2	Q55S44_CRYNE
35	132	14.3	611	2	Q412A2_GIBZE
36	131.5	14.2	326	2	Q22514_USEUD
37	131.5	14.2	763	2	Q9XDH2_MYCTU
38	131.5	14.2	1033	1	IF2_STRCO
39	131	14.2	297	2	Q527W5_MAGGR
40	131	14.2	498	2	Q5KJ55_CRYNE
41	131	14.2	3204	2	Q6X248_GALPH
42	130.5	14.1	708	2	Q9SX31_ARATH
43	130.5	14.1	1486	2	Q4VXV7_9DELT
44	130	14.1	191	2	Q82327_ARATH
45	130	14.1	191	2	Q9C550_ARATH

## ALIGNMENTS

RESULT 1	
Q8N7G8_HUMAN	
ID Q8N7G8_HUMAN PRELIMINARY;	PRT; 489 AA.
AC Q8N7G8	
DT 01-OCT-2002 (TEMBLrel. 22, Created)	
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)	
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)	
DE Hypothetical protein FLJ25605.	
GN Name=C14orf49;	
OS Homo sapiens (human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae;	
OC Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP NUCLEOTIDE SEQUENCE.	
RC TISSUE=Thyroid;	
RA Nishimura K., Magatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,	
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,	
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,	
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama A., Kawakami B.,	
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,	
RA Nagai K., Isogai T., Sugano S.,	
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.	
DR EMBL; AK098471; BAC05312.1; -; mRNA.	
DR Ensembl; ENSG00000176438; Homo sapiens.	
DR HGNC; HGNC:19861; C14orf49.	
DR InterPro; IPR002017; Spectrin.	
DR SMART; SM00150; SPEC; 1.	
SQ SEQUENCE 489 AA; 56176 MW; 19ABDBDE0AC6EEF CRC64;	

## Alignment Scores:

Pred. No.:	4,14e-38	Length:	489
Score:	592.00	Matches:	117
Percent Similarity:	95.2%	Conservative:	1
Best Local Similarity:	94.4%	Mismatches:	6
Query Match:	64.1%	Indels:	0
DB:	2	Gaps:	0

US-10-071-510A-16 (1-493) x Q8N7G8\_HUMAN (1-489)

```

QY 3 GCGGCGCCGGGGATGCGCAGATCCCAAGAGCCCGAGTTTGAGAGCGCTGGTGCAGAAATTC 62
DB 192 AAGlyProGlyAspAlaGluSerGlnGluAlaGlnPheGluValAlaGluPhe 211
QY 63 CCGGAGAAGAGAGCCCGAGCTGTCTCTGTGGAAGCCGAGGCGCTGGCTGATGAGAAG 122
DB 212 ProGlnValGlnAlaGlnIleuSerLeuValGlnAlaGlnGlyTyrPheValMetGlnVal 231
QY 123 TCTTCTCCGAGAGGCTGCTGCGCTGTGTCAGAGAGAGCTCAGGAGCTGGCAGAGTGTGG 182
DB 232 SerSerProGlnGlyAlaAlaValAlaGlnGlnGluValArgGlnIleuValGluSerTyr 251
QY 183 CCGGCGCTTGAGGCTGTGGAAGAAAGTCTGAGCTCATCGAAGAACTGGCATTCGCGAG 242
DB 252 ArgAlaLeuArgLeuLeuGlnGluSerLeuLeuSerLeuIleArgAsnTyrPheGln 271
QY 243 AGGATGGAAGTGAATTCGGGGAAGAAATGCTTTTCACCAACAATCCCAAGTCAGGA 302
DB 272 ArgMetGlnValAlaAspSerGlyValMetValPheThrAsnIleProlysserGly 291
QY 303 TTTCATCAATCCATGATCTCTATTCGAGGATCGTCGAGCGCTGAGTCTGTACG 362
DB 292 PheLeuIleAsnProMetAspProIleProArgHisArgArgAlaAsnLeuGln 311
QY 363 AGGCGTGTGGA 374
DB 312 GlnGlnGlnGly 315

RESULT 2
06ZM23_HUMAN PRELIMINARY; PRT; 904 AA.
ID 06ZM23_HUMAN
AC 06ZM23
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ16564.
GN Name=C14orf49;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagaetsuma M., Murakami B., Suzuki Y., Sugano S., Nagahara K.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
RA Masuhio Y., Nagai K., Isogai T.,
RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131436; BAD18582.1; -; mRNA.
DR HGNC; HGNC:19661; C14orf49;
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00435; Spectrin; 2.
DR SMART; SM00150; SPEC; 2.
SQ SEQUENCE 904 AA; 104056 MW; CAAC7405B17CAC9 CRC64;

Alignment Scores:
Pred. No.: 1,34e-37 Length: 904
Score: 586.00 Matches: 115
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 63.5% Indels: 0
DB: 2 Gaps: 0

US-10-071-510A-16 (1-493) x 06ZM23_HUMAN (1-904)
QY 3 GCGGCGCCGGGGATGCGCAGATCCCAAGAGCCCGAGTTTGAGAGCGCTGGTGCAGAAATTC 62

```

```

DB 612 AlaGlyProGlyAspAlaGluSerGlnGluAlaGlnPheGluValAlaGluPhe 631
QY 63 CCGGAGAAGAGAGCCCGAGCTGTCTCTGTGGAAGCCGAGGCGCTGGCTGATGAGAAG 122
DB 632 ProGlnValGlnAlaGlnIleuSerLeuValGlnAlaGlnGlyTyrPheValMetGlnVal 651
QY 123 TCTTCTCCGAGAGGCTGCTGCGCTGTGTCAGAGAGAGCTCAGGAGCTGGCAGAGTGTGG 182
DB 652 SerSerProGlnGlyAlaAlaValAlaGlnGlnGluValArgGlnIleuValGluSerTyr 671
QY 183 CCGGCGCTTGAGGCTGTGGAAGAAAGTCTGAGCTCATCGAAGAACTGGCATTCGCGAG 242
DB 672 ArgAlaLeuArgLeuLeuGlnGluSerLeuLeuSerLeuIleArgAsnTyrPheGln 691
QY 243 AGGATGGAAGTGAATTCGGGGAAGAAATGCTTTTCACCAACAATCCCAAGTCAGGA 302
DB 692 ArgMetGlnValAlaAspSerGlyValMetValPheThrAsnIleProlysserGly 711
QY 303 TTTCATCAATCCATGATCTCTATTCGAGGATCGTCGAGCGCTGAGTCTGTACG 347
DB 712 PheLeuIleAsnProMetAspProIleProArgHisArgArgArg 726

RESULT 3
08C117_MOUSE PRELIMINARY; PRT; 895 AA.
ID 08C117_MOUSE
AC 08C117
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
DE library, clone:4831426119 product:hypothetical Spectrin repeat
DE containing protein, full insert sequence.
GN Name=4831426119Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Queckenbush J.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Guscinich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombere P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsunki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The FANTOM Consortium,

```

the RIKEN Genome Exploration Research Group Phase I & II Team;  
"Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
Nature 420:563-573 (2002).  
(4)  
NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carinci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
Genome Res. 10:1617-1630(2000).  
(5)  
NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carinci P.,  
RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Sumi N., Iehi Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
RA Yoneba Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384 format  
sequencing pipeline with 384 multicapillary sequencer.";  
Genome Res. 10:1757-1771(2000).  
(6)  
NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carinci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Iehi Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata H., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ono M., Onatsu N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takawa-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK029216; BA026351.1; -; mRNA.  
DR Ensembl; ENSMUSG0000054150; Mus musculus.  
DR MGI; MGI:2442408; 483142619gk.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF00435; Spectrin; 2.  
DR SMART; SM00150; Spec; 2.  
DR Hypothetical protein.  
SQ SEQUENCE 895 AA; 101521 MW; BFF3E42537DA0C17 CRC64;

Alignment Scores:  
Pred. No.: 3,13e-23 Length: 895  
Score: 404.00 Matches: 90  
Percent Similarity: 71.1% Conservative: 23  
Best Local Similarity: 56.6% Mismatches: 36  
Query Match: 43.8% Indels: 10  
DB: 2 Gaps: 4

US-10-071-510a-16 (1-493) x Q8C17\_MOUSE (1-895)

QY 15 GATGCGGAGTCCCAAGAGCGGAGTTTGAAGAGCTGTGGCAGAGATTCGCGAGAGAGAG 74  
DB 594 AepHlaeIuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 613  
QY 75 GCCCAGCTGTCTCTGTGGAAGCGCAGAGGCTGCTGTGATGAGGAAGTCTTCTCCGAG 134  
DB 614 ValGlnValSerLeuLeuGlnAlaLeuGlnGlnLeuValMetLeuValSerSerProGlu 633  
QY 135 GGTGCTCCGCTGTGTCGAG 194  
DB 634 G1yAlaThrMetValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 653

QY 195 CTTGCGAAGAAAGTCTGTGAGCCTCATCAGAACTGGCATCTGCAGAGATGAGAGT 254  
DB 654 LeuLeuGlnGlnValSerMetLeuSerLeuMetArgAsnGlnGlnGlnGlnGlnGlnGln 673  
QY 255 GATTCGGGAGAGAAATGTTTTCACCAACATCCCAAGTCCAGAGATTTCTCATAT 314  
DB 674 AepHnGlyLysGlnGlnValPheThrAsnMetIleProLysAlaGlyPheLeuIleAsn 693  
QY 315 CCCATGATCTTATTTCCAGGAGCATCTGCAGCCGTCAGT---CTGTCTAGCAGGCTGTG 371  
DB 694 ProGlnAspProIleProArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 713  
QY 372 GGAGAAGGGCCAGGCCCATGCAAGAG-----TGGGTAGGCGCTTCCACG 419  
DB 714 ProGlnAsp---HisProGlnLeuLeuArgAspPheGlnGlnGlnGlnGlnGlnGlnGln 732  
QY 420 ACAGGC-----CCCTCCCTGTCGTGGGCAACATGCTGCTGTGAGACT 464  
DB 733 SerLysLeuArgArgIleIleThrMetArgValAlaThrAlaLysAspLeuArgThr 751

RESULT 4  
Q4FZC9\_MOUSE PRELIMINARY; PRT; 975 AA.  
ID Q4FZC9\_MOUSE PRELIMINARY; PRT; 975 AA.  
AC Q4FZC9\_MOUSE PRELIMINARY; PRT; 975 AA.  
DT 13-SEP-2005 (TRENDEL, 31, Created)  
DT 13-SEP-2005 (TRENDEL, 31, Last sequence update)  
DT 13-SEP-2005 (TRENDEL, 31, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
NCBI TaxID=10090;  
[1]  
NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eye;  
RX MEDLINE=22288257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins L., Wagner L., Shenmen C.F., Bhat N.K.,  
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Toehyuk S., Carinci P., Mulvaney S.J.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulvaney S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S.N., Krzywiński M.I., Skalska U., Smalins D.E.,  
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
NUCLEOTIDE SEQUENCE.  
RC TISSUE=Eye;  
RG NIH MGC Project;  
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC099694; AA099694.1; -; mRNA.  
DR Hypothetical protein.  
SQ SEQUENCE 975 AA; 112035 MW; 26D80A295CE0CF8B CRC64;

Alignment Scores:  
Pred. No.: 3,17e-23 Length: 975  
Score: 404.00 Matches: 90  
Percent Similarity: 71.1% Conservative: 23  
Best Local Similarity: 56.6% Mismatches: 36  
Query Match: 43.8% Indels: 10  
DB: 2 Gaps: 4





```

Db      1 MetLysSerSerProGluGlyAlaThrMetValGlnGluGluLeuArgLysLeuMet 20
QY      174 GAGTGTGGCGGGCCCTTGAGGCTGGAGAAAGTCTGAGCCTCATGCAAACTGG 233
Db      21 GUSERTGPGLAlaLeuArgLeuGluGluMetLeuSerLeuMetArgLysGln 40
QY      234 CATCTGACAGAGATGAGAAATGATTGCGGGAGAAATGTTTCCACCAACATCCCA 293
Db      41 GlnLeuGlnArgThrGluValAspThrGlyLysGlnValPheThrAsnLeuPro 60
QY      294 AAGTACGATTTCTCATCATCCATGATCTCTATTCCAGGATGTCGAGCGGAGT 353
Db      61 LysAlaGlyPheLeuLeuAsnProGlnAspProLeuArgArgLleIleThrMetArgValAlaThr 80
QY      354 ---CTGTCTACGACGGCTCTGGAGAGGGCCAGGCTCCAGTCAAGAG----- 401
Db      81 ProLeuGluGlyHisAspLeuProGluAsp---HisProGlnLeuLeuArgAspPheGlu 99
QY      402 ---TGGGTAGGGGTCTCCAGACAGAGC-----CCTCTCTGTCTGGGCGACAA 446
Db      100 GlnTProLeuGlnAlaLeuAsnSerLysLeuArgArgLleIleThrMetArgValAlaThr 119
QY      447 TGCTGTGCTGTGAGAGACT 464
Db      120 AlaLysAspLeuArgThr 125

RESULT 6
GPI_CHLRE
ID      GPI_CHLRE STANDARD; PRT; 555 AA.
AC      Q9FP06; Q03927;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      10-MAY-2005 (Rel. 47, Last annotation update)
DE      Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
GN      Name=GPI;
OS      Chlamydomonas reinhardtii.
OC      Chlamydomonadales; Chlorophyta; Chlorophyceae;
OC      Chlamydomonadales; Chlamydomonadaceae; Chlamydomonadaceae.
OX      NCBI_TaxID=3055;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RX      MEDLINE=21159092; PubMed=11258910; DOI=10.1021/bi0023605;
RA      Ferris P.J., Moessner J.P., Waffenschmidt S., Kiltz S., Drees J.,
RT      Goodenough U.W.;
RL      "Glycosylated polyproline II rods-with-kinks as a structural motif in
      plant hydroxyproline-rich glycoproteins.";
      Biochemistry 40:2978-2987(2001).
RN      [2]
RP      PRELIMINARY PARTIAL NUCLEOTIDE SEQUENCE.
RX      MEDLINE=91017504; PubMed=1699225;
RA      Adair W.S., Apt K.E.;
RT      "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
      encoding cell wall hydroxyproline-rich glycoproteins.";
      Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
RL      -1- FUNCTION: Major component of the outer cell wall W6 (crystalline)
      layer.
      -1- SUBUNIT: Associates with GPI2 and GPI3.
      -1- PTM: N-glycosylated and O-glycosylated.
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use as long as its content is in no way modified and this statement is not
      removed.
CC      EMBL: AF309494; AAC5420.1; -; Genomic DNA.
CC      EMBL: MS8496; AAC69706.1; ALT_SEQ; mRNA.
CC      GlycosylatedDB: Q9FP06;
DR      InterPro: IPR003882; Pfam1_extensin.
DR      PRINTS: PR01218; PSTLEXTENSIN.
KW      Glycoprotein; Repeat; Signal.

```

```

FT      SIGNAL 1 29 Potential.
FT      CHAIN 30 555 Vegetative cell wall protein gpi.
FT      --REGION 40 339 49 X 5 AA approximate repeats of P-P-S-P-
FT      COMPBIAS 259 279 X.
FT      CARBOHYD 399 399 Poly-Pro.
FT      CARBOHYD 455 455 N-linked (GlcNAc...) (Potential).
FT      CARBOHYD 493 493 N-linked (GlcNAc...) (Potential).
SQ      SEQUENCE 555 AA; 54220 MW; 6A584A90465502F5 CRC64;

Alignment Scores:
Pred. No.: 0.00397 Length: 555
Score: 149.00 Matches: 53
Percent Similarity: 36.88 Conservative: 3
Best Local Similarity: 34.98 Mismatches: 66
Query Match: 16.1% Indels: 31
DB: Gaps: 6

US-10-071-510A-16 (1-493) x GPI_CHLRE (1-555)
QY      446 TGTGTCGAGAGACCCCTTACCCACTCTTGAC 393
Db      31 CysValPro-----GlyGlyIlePheAsnCySProProSerProAlaProProSerPro 48
QY      392 CTGGGGCCCTGGCCCTTCTCCACAGCCCTGTAGACAGACTCAGCGCTGACGATGCTT 333
Db      49 AlaProProSerProAlaProProSerProAlaProProSerProAlaProProSerPro 68
QY      332 GCGAATAG-GATCCATGGGATGATGAGAAATCTGACTTGGATGTTGTTGTAAGA 274
Db      69 GlyProProSerPro-----Ala 74
QY      273 CCATTTCTTCCCGCATCTTCCATCTCTGCAATGCCAGTTTCTGATGAGGCTCA 214
Db      75 ProProSerProProSerProAlaProProSer-----Pro 86
QY      213 GCAGACTTCTTCCAGACCCCTCAAGCGCCGACAGACTCTGCGACTCCCTGAGCTCT 154
Db      87 AlaProProSerProAlaProProSerProAlaProProSerProAlaProProSerPro 106
QY      153 CTGACACAGCGAGACCCCTCGGAGAAAGACTTTCATCAGACAGACCCCTGCGCTT 94
Db      107 AlaProProSerProAlaProProSerProAlaProProSerProAlaProProSerPro 126
QY      93 CCA---CAAGGACAGCTGGGCTCTTCTCGGGAATTTGCCACAGACTCTCAACT 37
Db      127 ProSerProSerProProAlaProProSerProSer-----ProProSerProAla 143
QY      36 CGGCTCTTGGGACTGGCATCCCGGGCGCG 1
Db      144 ProProLeuProProSerProAlaProProSerPro 155

RESULT 7
QASMF7_TETNG
ID      QASMF7_TETNG PRELIMINARY; PRT; 911 AA.
AC      QASMF7;
DT      13-SEP-2005 (TrEMBLrel. 31, Created)
DT      13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT      13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE      Chromosome undetermined SCAP14551, whole genome shotgun sequence.
GN      ORFNames=GSTBN00015781001;
OS      Tetradon nigriviridis (Green puffer).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC      Acanthomorpha; Acanthopterygii; Percomorphia; Tetraodontiformes;
OC      Tetraodontidae; Tetraodon.
OX      NCBI_TaxID=99883;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RX      Nucleotide sequence.
RA      Mancel E., Bouneau L., Fischer C., Ozouf-Coetzee C., Bernot A.,
      Nicand S., Jaffe D., Fisher S., Lutfalla G., Dosset C., Segurens B.,
      Dasilva C., Salenouat M., Levy M., Boudet N., Castellano S.,

```



Db 1377 ProSerProGlnProProSerProSerProSerProAlaProAlaProProSerProAlaProLeuGlu 1396  
Qy 51 CCAGCCTCTCAAACTCGGCTCTTGAGACTCGGCAATCCCGGCGCG 1  
Db 1397 ProAlaAlaProAlaProProGlyProProProGlnProProGlyAlaPro 1413

RESULT 9  
Q948Y6 VOLCA  
ID Q948Y6 VOLCA PRELIMINARY; PRT; 1143 AA.  
AC Q948Y6:  
DT 01-DEC-2001 (TRENBLREL. 19, Created)  
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)  
DE VMP4 protein.  
GN Name=vmp4;  
OS Volvox carterii f. nagariensis.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
OC Chlamydomonadales; Volvocaceae; Volvox.  
OX NCBI\_TaxID=3068;  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=HK 10;  
RX MEDLINE=21382906; PubMed=11489172;  
RA Hallmann A.; Amon P.; Godl K.; Helitzer M.; Sumper M.;  
RT "Transcriptional activation by the sexual pheromone and wounding: a  
RT new gene family from Volvox encoding modular proteins with  
RT (hydroxy)proline-rich and metalloprotease homology domains."  
RL Plant J. 26:583-593(2001).  
EMBL: AJ311550; CAC39319.1; -; mRNA.  
DR MEROPS; M11.002;  
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.  
DR GO; GO:0005529; F:sugar binding; IEA.  
DR InterPro; IPR003014; PAN.  
DR InterPro; IPR003609; Pan\_app.  
DR InterPro; IPR008752; Peptidase\_M11.  
DR InterPro; IPR003882; Pileil\_extensin.  
DR InterPro; IPR00772; Ricin\_B\_lectin.  
DR Pfam; PF00024; PAN; 1.  
DR Pfam; PF05548; Peptidase\_M11; 1.  
DR Pfam; PF06652; Ricin\_B\_lectin; 2.  
DR PRINTS; PRO1218; PSTLEXTENSIN.  
DR PROSITE; PS50948; PAN; 2.  
DR PROSITE; PS50231; RICIN\_B\_LLECTIN; 2.  
SQ SEQUENCE 1143 AA; 123155 MW; 0CE0FD2CED52200C CRC64;

Alignment Scores:  
Pred. No.: 0.0188 Length: 1143  
Score: 141.00 Matches: 42  
Percent Similarity: 32.3% Conservative: 10  
Best Local Similarity: 26.1% Mismatches: 55  
Query Match: 15.3% Indels: 54  
Gaps: 4

US-10-071-510A-16 (1-493) x Q948Y6\_VOLCA (1-1143)  
Qy 483 ATGAGCAGAGCGGCGCAATGCTCTGAGCAGACGATGTCGCCAGAGGAGGCG 424  
Db 486 MetleuGluGlyIleAlaValIySthSerProProheValProProProProSerPro 505  
Qy 423 CTGTGTGAGACCCCTTACCTCTTGACCTGGGGCTGGCCCTTTCCACAGCCC 364  
Db 506 LeuIeuThSerProArgProProSer-----ProArgProProArgPro 520  
Qy 363 TGTCTAGACAGCTACCGCTGACGATGCTGGGATATGATCATGGATTGATGAA 304  
Db 520 ----- 520  
Qy 303 ATCTGACTTTGGATGTGTGTGTAAGAAACATTTCTTCCCGAATTCATTCATCC 244  
Db 521 -----SerProProSerProProPro----- 527  
Qy 243 TGTGAGATGCGAGTTTCTGTATGAGGCTCAGCAGACTTTCTTCAGACGCTCAAGGCC 184

Db 528 -----ProProSerProProPro 533  
Qy 183 GCCAGACTTCCAGCTTCCTGAGCTCTCTGACACGACGACAGACCCCTCGAGAG 124  
Db 534 ProProSerProProProProProSerProProProProSerProProProPro 553  
Qy 123 ACTTCTCATACGACCGCCGCTTCCACGAGGAGACAGTGGCCCTCTCTCCG 64  
Db 554 SerProProProProProProProProProProProProProProProProSerPro 573  
Qy 63 GGAATTCGACGACGCTCTCAAACTCGGCTCTTGAGACTCGGCAATCCCGGCGCG 4  
Db 574 Pro-----ProProProSerProProProProProProSerProArgHisProProSerPro 591  
Qy 3 CCG 1  
Db 592 Pro 592

RESULT 10  
Q6SSB8 CHLRE  
ID Q6SSB8 CHLRE PRELIMINARY; PRT; 3889 AA.  
AC Q6SSB8;  
DT 05-JUL-2004 (TRENBLREL. 27, Created)  
DT 05-JUL-2004 (TRENBLREL. 27, Last sequence update)  
DT 05-JUL-2004 (TRENBLREL. 27, Last annotation update)  
DE Minus agglutinin.  
GN Name=SAD1;  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=CC-621;  
RX PubMed=15659633; DOI=10.1105/tpc.104.028035;  
RA Ferris P.J.; Walfenschmidt S.; Umen U.G.; Lin H.; Lee J.H.; Ishida K.;  
RT "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii."  
RL Plant Cell 17:597-615(2005).  
EMBL: AY450929; AAS07042.1; -; Genomic DNA.  
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.  
DR InterPro; IPR002951; Atrophin.  
DR InterPro; IPR006315; Autotransporter.  
DR InterPro; IPR003882; Pileil\_extensin.  
DR PRINTS; PRO1222; ATROPHIN.  
DR PRINTS; PRO1218; PSTLEXTENSIN.  
DR TIGRFAMs; TIGR01414; autotrans\_bart1; 1.  
SQ SEQUENCE 3889 AA; 389223 MW; 4DE5B44D5507214A CRC64;

Alignment Scores:  
Pred. No.: 0.0243 Length: 3889  
Score: 140.50 Matches: 43  
Percent Similarity: 33.1% Conservative: 4  
Best Local Similarity: 30.3% Mismatches: 42  
Query Match: 15.2% Indels: 53  
Gaps: 5

US-10-071-510A-16 (1-493) x Q6SSB8\_CHLRE (1-3889)  
Qy 411 CCCTTACCCAGCTTGTGACTGGGCGCTGCTCTCCACAGCCCTGCTAGACAGAC 352  
Db 1369 ProProProProSer-----ProThProProSerPro----- 1379  
Qy 351 TCACGGGTGACGATGCTGGGAATGATGATGATGATGATGATGATGATGATGATG 292  
Db 1379 ----- 1379  
Qy 291 GATGTTGTTGTAAGAAACATTTCTTCCCGAATTCATTCATTCATTCATTCATTC 222  
Db 1380 -----GlnProProValProProSerProThProProSer----- 1391  
Qy 231 AGTTTCTGATGAGGCTCAGAGACTTTCTTCAGCAGGCTCAAGCCGCGCAGACTCTG 172

```
DB 1392 -----ProGlnProProserProAlaProProser 1401
QY 171 CCAGCTCCTGAGCTCTCTCGACACGCGACACCTTCGGAGAAGATTTCATCA 112
DB 1402 ProAlaProSerAlaProLeuGlnProSerProAspProProserProGlnProProser 1421
QY 111 CCAGCC-----AGCCCTGGGCTTCACCGACGAGCTGGGCTCTCTCT 67
DB 1422 ProAlaProGlnProProProserProProProserProSerThrProSerProProser 1441
QY 66 CCGGGAATTCTGCGACGAGCTCTCAAACTCGGCTTTGGGACTCGGCATCCCGCGGC 7
DB 1442 ProAlaProLeuAlaProAlaProProValProProMetAlaProGln---ProProser 1460
QY 6 CCGCCG 1
DB 1461 ProPro 1462

RESULT 11
Q6SSB6_CHLRE PRELIMINARY; PRT; 3409 AA.
ID Q6SSB6_CHLRE
AC Q6SSE6;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Plus agglutinin.
GN Name=SAG1;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxId=3055;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15659633; DOI=10.1105/epc.104.028035;
RX Ferris P.J., Waffenschmidt S., Umen J.G., Lin H., Lee J.H., Ishida K.,
RX Kubo T., Lau J., Goodenough U.W.;
RT "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RT Plant Cell 17:597-615(2005).
DR EMBL; AY450930; AAS0704.1; -; Genomic DNA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pstc11_extensin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01218; PST1EXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633ED6631F CRC64;

Alignment Scores:
Pred. No.: 0.0286 Length: 3409
Score: 139.50 Matches: 53
Percent Similarity: 37.4% Conservative: 14
Best Local Similarity: 29.6% Mismatches: 57
Query Match: 15.1% Indels: 55
DB: 2 Gaps: 7

US-10-071-510A-16 (1-493) x Q6SSB6_CHLRE (1-3409)
QY 467 CCAAGTCTCAAGACGACGATGTTCCCCACAGAGGAGGGCTGTGCTGAGACCCC 408
DB 505 ProserProProserProProserProPro-----Pro 515
QY 407 TACCCACCTTGTGACCTGGGGCTGGCCCTTCTCCACAGCCCTGTAGACAGACTAC 348
DB 516 SerProProGlnProProserProProProLeuProProserProProserProThrPro 535
QY 347 GCGTCGACGATCCCTGGGATAGATCCATGGGATTGATGAGAAATCTGACTTTGGAT 288
DB 536 ValAla-AlaArgCysIleGlnValGly-----GlyT1 545
QY 287 GTTGTGTGTAACCAATTTCTTCCCGAATCCACTTCCATCTCT----- 241
DB 545 e-Cys-----AspSerProserProMetProProserProArgProProG 560
QY 240 -----GCAGATG 234
```

```
DB 560 InProProserProProProProProProArgProProProAlaProArgProserP 580
QY 233 CCAAGTTTCTGATGAGGCTCAGACACTTTCTTCAGACGACTCAAGCCGCCAGACTC 174
DB 580 roProBehisProProserProAspSerProProAlaSerSerValPro----- 596
QY 173 TGCAGCTCCCTGAGCTCTCTCGACACGCGACACCCCTCGGAGAAGATTCTGCAT 114
DB 597 --ProserPro--GlnProProserProLys-----ProProserProAlaProPro 613
QY 113 CACACGACCCCTCGCTCGCTTCACACGAGGAGACTGGGCTCTCTCTCGGGAATTCTGC 54
DB 613 erProAlaProProserProProProProserProAlaProProserProAlaProPro 633
QY 53 CACGAGCTCTCAAACTCGGCTCTTGGGACTCGGCTCCCGGCGCGCG 1
DB 633 erProAlaProProserProGlnProProserProValProProGlnProPro 650

RESULT 12
Q58NA5_CHLIN PRELIMINARY; PRT; 2371 AA.
ID Q58NA5_CHLIN
AC Q58NA5;
DT 10-MAY-2005 (TREMBlrel. 30, Created)
DT 10-MAY-2005 (TREMBlrel. 30, Last sequence update)
DT 10-MAY-2005 (TREMBlrel. 30, Last annotation update)
DE Plus agglutinin (Fragment).
GN Name=SAG1;
OS Chlamydomonas incerta.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxId=51695;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=CC-1870;
RX Lee J.-H., Ferris P.J., Waffenschmidt S., Ishida K., Goodenough U.W.;
RT "Evolution of Genes Encoding Chimeric Hydroxyproline-Rich
RT Glycoproteins and Sex-Related Proteins in Two Species of
RT Chlamydomonas.";
RT Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY937239; AAX33674.1; -; Genomic DNA.
DR NON TER 2371
FT SEQUENCE 2371 AA; 236526 MW; A6E5F15927378BA CRC64;

Alignment Scores:
Pred. No.: 0.0298 Length: 2371
Score: 139.00 Matches: 50
Percent Similarity: 35.4% Conservative: 6
Best Local Similarity: 31.6% Mismatches: 47
Query Match: 15.1% Indels: 56
DB: 2 Gaps: 6

US-10-071-510A-16 (1-493) x Q58NA5_CHLIN (1-2371)
QY 473 ACGTGCCCAAGTCCACAGACGATGTTCCCCACAGAGGAGGGCTGTGCTGGA 414
DB 1215 ThrProProAlaProGlnProProserProAlaProProAla----- 1228
QY 413 GACCCCTTCAACCTTGTGACCTGGGGCTGGCCCTTCTCCACAGCCCTGTAGACAG 354
DB 1229 --ProGlnProProserProAlaProProserProAsnProProserProAlaProThr 1247
QY 353 ACTCAGCGGTGACGATGCTGGGAATAG-GATCCATGGGATTGATGAAATCTGACT 295
DB 1248 ThrProAlaSerProGlnProProserProGlnPro----- 1259
QY 294 TTGGATGTGTGTGTAACCAATTTCTTCCCGAATCCACTTCCATCTGCAAT 235
DB 1260 -----ProserProserPro----- 1264
QY 234 GCGAGTTTCTGATGAGGCTCAGACACTTTCTTCAGACGACTCAAGCCGCCAGACT 175
DB 1265 -----ProValProProserProAlaProPro 1273
```

```
QY 174 CTGCCAGCTCCCTGAGCTCTCTGACACGAGCAGACCTTCGAGAGAATTCTTCCA 115
    |||||
DB 1274 SerProAlaProLeuProProSerPro-----AspProPro-----SerPro 1288
QY 114 TCACCAAGCAGCCCTGCGCTTCCACGAGGACAGCTGGGCTCTTCTCCGGAATTCTG 55
    |||||
DB 1289 ValProProSerProAlaProProProSerProProAlaProProSerProGluProIle 1308
QY 54 CCACCAAGCTCTCAAACTCGGCTCTTGGGACTCGGCATCCCGGCGCGCG 1
    |||||
DB 1309 ProProAlaPro-----ProProSerProPro 1317

RESULT 13
Q7SF15_NEUCR
ID 07SF15_NEUCR PRELIMINARY; PRT; 636 AA.
AC 07SF15;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU07438.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehm B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Seltremikoff C.P., Kinsey U.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gaele S.,
RA Kamal M., Kamysselet M., Mauceli E., Bielke C., Rudd S., Fishman D.,
RA Kryofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cognoli C., Macino G., Catchside D., Ij W., Pratt R.J., Osmann S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voecker R.,
RA Yarden O., Plamann M., Selter S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbold D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nisbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0-0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -!- SUBUNIT: Isoform 1 and isoform 2 encode coiled-coil structures
CC that mediate homo- and heteromultimerization (By similarity).
CC -!- DOMAIN: The WH1 domain interacts with the PPXX motif in GRM1,
CC GRM5, RYR1, RYR2, ITPR1, SHANK 1 and SHANK3 (By similarity).
CC EMBL; AABX0100025; EAA3589.1; -; Genomic_DNA.
DR InterPro: IPR000697; PH1.
DR InterPro: IPR011993; PH type.
DR InterPro: IPR001960; WH1.
DR InterPro: IPR00124; WH2.
DR Pfam; Pf00568; WH1; 1.
DR Pfam; Pf0205; WH2; 1.
DR KEGG; K01205; WH2; 1.
KM Coiled coil; Hypothetical protein.
SQ SEQUENCE 636 AA; 66201 MW; AE102B813CA938B6 CRC64;
```

```
DB 407 ProLeuProProLysAlaPro-----GlyProAlaProProLeuProPro 421
QY 198 GCAGCTCAAGAGCCCGCCAGCAGACTGCGACGCTCCCTGACCTCTGACAGGAG 139
    |||||
DB 422 AlaSerSerArgProProProMetLeuProThr-----ArgSerProAlaProProGln 439
QY 138 CACCTCCGAGAGAACTTCTTCATCACCAGCAGCCCTGCGCTTCCA----- 91
    |||||
DB 440 AlaProProLeuProThrSerAlaAlaProProProProProAlaThrGlnAla 459
QY 90 -----CAAGAGACAGCTGGGCTCTTCTCCGGAATTCTGCACCA 49
    |||||
DB 460 ProProProProProProLeuProAlaThrSerAlaProProProProProProAlaProPro 479
QY 48 GCCTCTCAAACTCGGCTCTTGGGACTCGGACTCCCGGCGCGCG 1
    |||||
DB 480 AlaProProAlaProProLeuProAlaAlaHisAlaProProProPro 495

RESULT 14
Q95JC9_PIG
ID 095JC9_PIG PRELIMINARY; PRT; 676 AA.
AC 095JC9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Basic proline-rich protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suidae;
OC Sus.
NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Parotid gland;
RX PubMed=15805110;
RA Zhang Q., Szalay A.A., Tietche J.M., Kye-Yune-Nyombi E., Sands J.F.,
RA Oberg K.C., Leonora J.;
RT "Cloning and Functional Study of Porcine Parotid Hormone, a Novel
RT Proline-rich Protein.";
RL J. Biol. Chem. 280:22233-22244(2005).
DR EMBL; AY035849; AAK61383.1; -; mRNA.
DR GO; GO:0016021; C:intracellular to membrane; IEA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR GO; GO:0005249; F:voltage-gated potassium channel activity; IEA.
DR GO; GO:0006813; P:potassium ion transport; IEA.
DR InterPro: IPR002951; Atrophin.
DR InterPro: IPR005404; KVs_3 channel.
DR InterPro: IPR003882; Pstc1L_extensin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01582; KV33CHANNEL.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 676 AA; 62297 MW; 3008BC41EPD81FC9 CRC64;
```

```
Alignment Scores:
Pred. No.: 0.0301 Length: 676
Score: 138.00 Matches: 60
Percent Similarity: 36.4% Conservative: 3
Best Local Similarity: 34.7% Mismatches: 58
Query Match: 15.0% Indels: 52
DB: 2 Gaps: 11

US-10-071-510A-16 (1-493) x Q95JC9_PIG (1-676)
```

```
QY 467 CCAAGCTCTGAGAGAGATGTCGCCAGAGAGAGGCTTGCTGAGAGAC--- 411
    |||||
DB 417 ProProProProProProProAlaAlaGluProGlnGlnProAlaProSerGlyAspLys 436
QY 410 -----CCCTACCAAGCTCTTGAACCTGGG-----CTGAGCCCTTCTCCACAGC 366
    |||||
DB 437 ProLysLysLysProProProProAlaGlyProProProProGlyProProSerProGly 456
QY 365 CTGCTTGAAGAGACTCAGCGCTGAGAGATGCTGGAGAAATGAGATTCATGGAGATTGATGAG 306
```

US-10-071-510A-16 (1-493) x Q7SF15\_NEUCR (1-636)

US-10-071-510A-16 (1-493) x Q95JC9\_PIG (1-676)

255 CCACTTCATCTCTGAGATGACGATTTCTGATGAGGCTCAGAGACTTTCTT---CCA 199

```

Db      457 ProAla-----ProProGlyAlaArg--Pro----- 464
Qy      305 AAATCTGACTTTGGATGTTGTGTAAMACCATTTTCTTCCCGAATCCACTTCAT 246
Db      465 -----ProProGlyProProProGlyProP 474
Qy      245 CCTCTGCAGATGCCATTTTCTGATGAGGCTCAGCAGACTTCTTCCAGCAGCCTCAAGC 186
Db      474 roProPro-----GlyProAla-----ProProGlyAlaArgP 485
Qy      185 CCGCCAGCACTCTCCAGCTCCC-----TGAGTCTCTCTGCACAC 144
Db      485 roProProGlyProProProProGlyProProProProGlyProAlaProProGlyAla 505
Qy      143 GCGACGACCTT---CCGAGAGAGACTTCTTCATCACCAGCCAGCCTGGCTTCCACAG 87
Db      505 rgProProProGlyProProProProGlyProProProProGlyProAla---ProProG 524
Qy      86 GGACA-----GCTGGGCTCTTCTCCGGAATTCTGCACAGCCTCTCAACTC 36
Db      524 lYAlaArgProProProGlyProProProProGly-----ProProProProGlyPro 542
Qy      35 GGCCTCTTGGACTCCGACATCCCGCGGCGCGCG 1
Db      542 laProProGlyAlaArgProProProProGlyProPro 553

RESULT 15
Q89X06 BRAJA PRELIMINARY; PRT; 745 AA.
AC Q89X06;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DE 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE B1r0521 protein.
GN OrderedLocusNames=b1r0521;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=USDA 110;
RX MEDLINE=2484998; PubMed=1259275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Ref. 9:189-197(2002).
DR EMBL; BA000040; BAC45786.1; -; Genomic_DNA.
DR HSSP; P01176; 1OAP.
DR GO; GO:0009279; C:outer membrane (sensu Gram-negative Bacteria); IEA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR InterPro; IPR003882; P1c1l_extensin.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR ProDom; PDO00930; OmpA/MocB; 1.
KM Complete proteome.
SQ SEQUENCE 745 AA; 74544 MW; 155EDFCC74DBC6D CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.0401 Length: 745
Score: 136.50 Matches: 48
Percent Similarity: 35.5% Conservative: 12
Best Local Similarity: 28.4% Mismatches: 65
Query Match: 14.8% Indels: 43
DB: 2 Gaps: 6

```

US-10-071-510A-16 (1-493) x Q89X06\_BRAJA (1-745)

```

Qy      441 CCCAGAGAGGAGGGGCGCTGTGTGAGAGACCCCTACCCACTTGTGAGGGCTGG 382
Db      181 ProProAlaAGlyProAlaAlaArgProThrProAlaProThrAlaThrPro-----Thr 198
Qy      381 CCCCTTCTCCACAGCCCTGTAGACAGACTCAGCGCTGACGATGCTGGGAATAGAT 322
Db      199 ProValAlaProProProAlaAlaProThrAlaArgProGlySerPro----- 214
Qy      321 CCATGGATGTGAGAAATCTGACTTTGGATGTGTGTAAMACCATTTTCTTCC 262
Db      215 -----AlaProAlaAlaThr 219
Qy      261 CCGAATCCACTTCATCTCTGCA-----GATGCCAGTTTCTGA 223
Db      220 ProAlaProThrProThrProAlaProThrAlaThrProAlaProThrAlaThrProAla 239
Qy      222 TGAGGCTCAGCAGACTTCTTCCAGCAGCCTCAAGGCCGCCAGC----- 178
Db      240 ProGlySerThrProGlyAlaBProProAlaGlyArgProGlyAlaBProProGlyAla 259
Qy      177 -----ACTCTCCAGCTCCCTGAGCTCTCTCTGACACAGCGACACACCTCCGAG 127
Db      260 ArgProGlySerProProAlaAlaGlySerProProAlaProGlyAlaThrProAlaPro 279
Qy      126 AAGACTTCTCATCACACAGCCAGCCCTGCGCTTCA-----CCAGGAGCAGCT 79
Db      280 ThrThrThrProAlaProGlyGlyThrAlaThrProProSerGlyArgProGlyProAla 299
Qy      78 GGGCTCTTCTCCGGAATTCTGCACACAGCTCTCAAACTCGGCTCTT----- 28
Db      300 SerThrProAlaProGlyAlaAlaThrProAlaProThrAlaThrProAlaProGlyGly 319
Qy      27 GGGACTCGGATCCCGCGGCGCGCGCG 1
Db      320 AlaLeuThrProProProGlyArgPro 328

```

Search completed: April 29, 2006, 03:47:43  
 Job time : 182.5 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OW nucleic - protein search, using frame\_plus.n2p model

Run on: April 29, 2006, 03:48:00 | Search time 6.6 Seconds  
(without alignments)  
1235.125 Million cell updates/sec

Title: US-10-071-510A-16

Sequence: 1 cggcgccggccgggggatgcgc.....cctgtcctcattgagcgtgc 493

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=x1p  
-Q=/abses/ABSSWEB.spool/US10071510/runat\_28042006\_170851\_18792/app.query.fasta\_1  
-DB=Issued Patents AA -QFMT=faetan -SUFFIX=rai -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abes02p -USR=US10071510 @CGN 1.1 71 @runat\_28042006\_170851\_18792  
-DCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONCLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5 COMB.pcp:\*  
2: /cgn2\_6/ptodata/1/iaa/H COMB.pcp:\*  
3: /cgn2\_6/ptodata/1/iaa/H COMB.pcp:\*  
4: /cgn2\_6/ptodata/1/iaa/PCRTUS COMB.pcp:\*  
5: /cgn2\_6/ptodata/1/iaa/RE COMB.pcp:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pcp:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	130	14.1	420	US-09-902-540-13993
C 2	127	13.8	553	US-09-949-016-7961
C 3	126.5	13.7	324	US-09-949-016-7664
C 4	126	13.7	339	US-09-252-991A-21715
C 5	125.5	13.6	561	US-09-252-991A-22317
C 6	123	13.3	334	US-09-823-240A-2
C 7	122.5	13.3	802	US-09-823-240A-2
C 8	122	13.2	331	US-09-949-016-9978
C 9	120.5	13.1	581	US-09-477-962-107
C 10	120.5	13.1	935	US-08-818-112-143
C 11	120	13.0	174	US-08-818-112-143
C 12	120	13.0	174	US-08-818-111-138

C 13	120	13.0	174	US-09-056-556-143	Sequence 143, App
C 14	120	13.0	174	US-09-072-596-138	Sequence 138, App
C 15	120	13.0	174	US-09-072-967-143	Sequence 143, App
C 16	120	13.0	174	US-10-193-002-138	Sequence 138, App
C 17	120	13.0	174	US-10-084-843-143	Sequence 143, App
C 18	120	13.0	701	US-09-252-991A-24048	Sequence 24048, A
C 19	119	12.9	525	US-08-764-870-7	Sequence 7, Appl1
C 20	119	12.9	525	US-08-980-115-7	Sequence 7, Appl1
C 21	119	12.9	533	US-07-952-800-2	Sequence 2, Appl1
C 22	119	12.9	533	US-08-216-592A-1	Sequence 4, Appl1
C 23	119	12.9	577	US-09-949-016-11572	Sequence 11572, A
C 24	118	12.8	439	US-10-300-819B-21	Sequence 21, Appl1
C 25	118	12.8	439	US-09-252-991A-22113	Sequence 22113, A
C 26	118	12.8	511	US-09-107-433-2804	Sequence 2904, App
C 27	117	12.7	104	US-09-547-693-235	Sequence 235, App
C 28	116	12.6	180	US-09-904-615-78	Sequence 78, Appl1
C 29	116	12.6	180	US-10-054-988-78	Sequence 4, Appl1
C 30	115.5	12.5	214	US-08-217-327-4	Sequence 4, Appl1
C 31	115.5	12.5	356	US-09-252-991A-18206	Sequence 18206, A
C 32	115.5	12.5	445	US-09-252-991A-23505	Sequence 23505, A
C 33	115	12.5	267	US-08-301-162-16	Sequence 16, Appl1
C 34	115	12.5	267	US-09-461-240-16	Sequence 16, Appl1
C 35	115	12.5	267	US-09-968-927-16	Sequence 16, Appl1
C 36	115	12.5	398	US-09-303-064-54	Sequence 54, Appl1
C 37	115	12.5	398	US-09-086-503-54	Sequence 54, Appl1
C 38	114.5	12.4	105	US-09-547-693-230	Sequence 230, App
C 39	114	12.4	72	US-09-547-693-231	Sequence 231, App
C 40	114	12.4	386	US-09-461-774-10	Sequence 10, Appl1
C 41	113.5	12.3	503	US-09-599-287A-2	Sequence 2, Appl1
C 42	113.5	12.3	503	US-10-078-547-2	Sequence 2, Appl1
C 43	112.5	12.2	226	US-09-252-991A-29689	Sequence 29689, A
C 44	112.5	12.2	1958	US-07-945-283-2	Sequence 2, Appl1
C 45	112.5	12.2	8991	US-08-714-741-32	Sequence 32, Appl1

## ALIGNMENTS

RESULT 1  
US-09-902-540-13993  
; Sequence 13993, Application US/0902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses thereof  
; FILE REFERENCE: 38-10(1584)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ. ID NOS: 16825  
; SEQ ID NO 13993  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-13993

## Alignment Scores:

Pred. No.: 0.000442  
Score: 130.00  
Percent Similarity: 36.6%  
Best Local Similarity: 31.3%  
Query Match: 14.1%  
DB: 2  
Gaps: 7

US-10-071-510A-16 (1-493) x US-09-902-540-13993 (1-420)

Oy 276 AAACACTTCTCTCC-----CCG 259  
Db 105 LysProhegUstergInValLeuLeuAapLyValValSalLeuValGlyGlnLysSer 124

```
QY 258 AATCCATTCATCCTCTGCGAGATGCCAGTTCTGATGAGCTCAGCAGATTCTTCCA 199
||| :|||:|||||
Db 125 AaThMePrProAlaSerAla-----AlaThrGlnValAlaGHis 137
QY 198 GCAGCTTCAGAGCCGCCGCGACTCTGCCAGCTCC-----TGAGCTCTCTGCA 148
||| :|||:|||||
Db 138 AlaAlaPrProGlnPrProAlaAlaAla---ProAlaPrProValAlaAlaAlaPrProGly 156
QY 147 CCACGGCAGACCCCTCCGAGAGAACTTCTCCATCA----- 112
||| :|||:|||||
Db 157 AlaArgPrProAlaPrProPrProGlyAlaArgPrProGlyValPrProGlyValPrPro 176
QY 111 -----CCAGCCAGCCCTCGGCTTCCA 91
||| :|||:|||||
Db 177 ArgPrProGlyAlaAlaGlyValPrProPrProGlyAlaArgPrProGlyValPrPro 196
QY 90 CCAGGAGCAGCTGGGCTCTCTCTCCGGAATTCTGCCAGCAGCTCT-----CAACT 37
||| :|||:|||||
Db 197 PrProGlyMetAlaArgPrProPrProGlyPrProGlyMetPrProPrProGlyAlaPrProGlyAlaPrPro 216
QY 36 CGGCTCTTGAGGAGCTCGGATCCCGCGGC-----CGCGCG 1
||| :|||:|||||
Db 217 ArgPrProGlyPrProGlyMetPrProPrProGlyMetAlaArgPro 230

RESULT 2
US-09-949-016-7961
/ Sequence 7961, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 7961
/ LENGTH: 553
/ TYPE: PRN
/ ORGANISM: Human
US-09-949-016-7961

Alignment Scores:
Pred. No.: 0.000936 Length: 553
Score: 127.00 Matches: 67
Percent Similarity: 35.9% Conservative: 7
Best Local Similarity: 32.5% Mismatches: 77
Query Match: 13.8% Indels: 56
DB: 2 Gaps: 10

US-10-071-510A-16 (1-493) x US-09-949-016-7961 (1-553)
QY 467 CCAAGCTCCAGAGCAGAGCATGTCGCCAGCAGAGGA----- 429
||| :|||:|||||
Db 191 PrProGlyGlySerSerArgAlaAlaPrProGlnGlyValArgGlyValArgGlyValArgPrPro 210
QY 428 GGGGCTGTGCTGAGACCCCTTACCACTTGTGACCTGGGGCTGGC----- 381
||| :|||:|||||
Db 211 GAlaAlaValPrProGlyValSerArgPrProPrProGlyPrProAlaGlyPrProGlyValPrProPrPro 230
QY 380 CCGTTCTCCACAGCCCTGCTAGACAGACTCAGCGCTCGACGATGCCCTGGGA---ATAG 324
||| :|||:|||||
Db 231 PrProPh-----PrProAlaGlyGlnThr-PrProArgPrProPrProGlyPrProPrPro 247
QY 323 ATCCATGGGATTTGATGAAATCCCTGACTTGGGATGTTGGTGAATAACATTTTCTT 264
||| :|||:|||||
```

```
Db 247 YPrProPrProGlyPrProPrProGlyPrProPrProGlyGlnValLeuPro-PrProLeuAlaG 267
QY 263 CCGCG-----AATCCATTCGATCTCTGAGATGACGAGTTCTGATGAG 219
||| :|||:|||||
Db 267 YPrProPrProAlaArgGlyValSerArgPrProPrPro-PrProValLeuPrProGlyValPrPro 286
QY 218 GCTCAGCAGACTTCTTCCAGAGGCTTCAGAGCCGCCAGC-ACCTGCGAGCT----- 166
||| :|||:|||||
Db 287 GAlaGlnPrProPrProGlyPrProLeuPrProPrProGlyPrProPrProPrProValPrProGlyTyrGly 306
QY 165 ---CCCTAGACTCTCTCTGCGACAGCAGCAGACCTCCGGAGAACTTCTCCATCACA 109
||| :|||:|||||
Db 307 PrProPrProGlyPrProPrProPrProGlnGlnGlyPrProPrProPrProGlyPrProPrPro 326
QY 108 GCCACCCCT-----GGGCTTCACACGAGGAGCAGCT 79
||| :|||:|||||
Db 327 PrProArgPrProPrProGlyPrProLeuGlyPrProPrProLeuThrLeuAlaPrProPrProHisLeuPrPro 346
QY 78 GGGCTCTCTCTCCGGAATTCTGCCAGCAGCTCTCAAACTCGGCTCTT----- 28
||| :|||:|||||
Db 347 GYPrProPr-----PrProGlyAlaPrProPrProPrProAlaPrProHisValaPrProAlaPrProPrPro 365
QY 27 -----GGGACTCGGCAATCCCGC----- 10
||| :|||:|||||
Db 366 PrProThrAsnSerGlyMetPrProThrSerAspSerArgGlyPrProPrProThrAspPrPro 385
QY 9 ---GGCGCGCG 1
||| :|||:|||||
Db 386 TyrGlyValArgPro 389

RESULT 3
US-09-949-016-7664
/ Sequence 7664, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 7664
/ LENGTH: 324
/ TYPE: PRN
/ ORGANISM: Human
US-09-949-016-7664

Alignment Scores:
Pred. No.: 0.000883 Length: 324
Score: 126.50 Matches: 52
Percent Similarity: 38.8% Conservative: 16
Best Local Similarity: 31.5% Mismatches: 66
Query Match: 13.7% Indels: 35
DB: 2 Gaps: 9

US-10-071-510A-16 (1-493) x US-09-949-016-7664 (1-324)
QY 440 CCCAGACAGAGGGGCTGTGCTGAGACCCCTTACCACTTGTGACCTGGGCTGGC 381
||| :|||:|||||
Db 125 PrProPrProGlyPrProAlaPrProSerGlyPrProPrProGlyPrProPrProGlnLeuPrProPrPro 144
QY 380 CCGTTCTCCACAGCCCTGCTAGACAGACTCAGCGCTCGACGATGCCCTGGGAATAGATC 321
||| :|||:|||||
Db 145 PrProGlyValHisPrProPrProAlaPrProValHisPrPro-----PrProAlaSerGlyVal 161
```

```

QY 320 CAT-----GGATGATGAGAAATCTGACTTTGGAGTGTGTGTAAGC 273
Db 162 HtProProAlaProGlyValHisProProAla-----ProGlyValHis 176
QY 272 CATTTTCTCCCGGATTCAC-----TTCCATCTCTGCGAGTCCAGTT 228
Db 177 ProProAlaProGlyValHisProProThrsGlyValHisProProAlaProGlyVal 196
QY 227 TCTGATGAGGCTCAGCAGACTTTCTTCCAGCAGCTTCAGGCCGCCAGCTCTGCCAG 168
Db 197 HtProProAlaProGlyValHis--ProProAla-----ProGlyValHisProPro 214
QY 167 CTCTCTGAGCT---CTCTCTGACACGCGCAGACCTT-----C 132
Db 214 lAProGlyValHisProProAlaProGlyValHisProProSerAlaGlyValHisP 234
QY 131 CGAGAGAAAGATTCTTCATCAGCAGCCGCTGCGCTTCACAGGAGACAGTGGG---- 76
Db 234 roGlnAlaProGlyValHisProAlaAlaProAlaValHisProGlnAlaProGlyValH 254
QY 75 --CTCTCTCTCCGCGAATTCTGCCACAGCCTCTCAAACTCGGCTCTTGGGACTCGGC 18
Db 254 lAProProAlaProGlyValHisProGlnAlaProGlyValHisPro-----G 270
QY 17 ATCCCCCGGCG 7
Db 270 lAProProGly 273

RESULT 4
US-09-252-991A-21715
; Sequence 21715, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21715
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21715

Alignment Scores:
Pred. No.: 0.001 Length: 339
Score: 126.00 Matches: 64
Percent Similarity: 27.6% Conservative: 13
Best Local Similarity: 22.9% Mismatches: 58
Query Match: 13.7% Indels: 144
Gaps: 12
US-10-071-510A-16 (1-493) x US-09-252-991A-21715 (1-339)

QY 2 GGCCGGCCGGGGG-----TCCGAGTCTCCAGAGGC 34
Db 69 GtYtYrProGlyValHisPProProValHisGlyLeuCybAsgAlaAsgGly 88
QY 35 CGAGTTTGAAGGCTGGTGGCAGAAATTCCTCCGAGAGAGGC----- 76
Db 89 GtYtYrProGlyValHisGtYtYrProGlyValHisPProProAlaGlyGlyAspGln 108
QY 77 -----CCAGCTGTCTCTGAGAGCGCAGG----- 103
Db 109 ValAlaHisValAlaAlaAlaLeuGlyGlySerArgGlyGlyArgSerThrGly 128
QY 103 ----- 103

```

```

Db 129 ArgAlaAerProProProAlaGlyValArgThrValGlnGluGlyArgArgArgProArg 148
QY 104 -----CTGCTGTGATGAGAACTTTCTCCGAGGCTGC 139
Db 149 ArgThrAlaProGlyValArgProArgAlaGlyValGly-----GlyGlyCys 164
QY 140 TCCCTGTGTCAGAGAGACTCAGAGAGCTGGCAGAGCTGTGTGCGGCTTGAGGCTGCT 199
Db 165 SerArgGlyAlaGlyAlaGlnAlaAlaGlyAlaGlyValAlaAlaAspAlaGly 184
QY 200 GGAAGAGTCTGCTGAGCT----- 220
Db 185 AspGlyValArgAlaAlaAlaProArgArgProValArgGluProProGlyAspProArgGly 204
QY 221 -----CATCAGAACTCGCATCTCCAGAGAT----- 247
Db 205 ThrValHisProArgThrHisGlnArgSerAlaGlyAlaProGluValArgAlaGly 224
QY 247 ----- 247
Db 225 AlaLeuHisProValValHisProGlyGlyGlnAlaArgArgGlyGlyHisLeuHis 244
QY 248 -----GGAAGTGAATTCGGGAGAGAAATGTTTTCAC 280
Db 245 GlyAspProArgThrGlyGlnGlyAlaAspGlyGlyThrGlyAlaGlnArgGlyPhe---- 263
QY 281 CAACAACATCCCAAGTCCAGATTCTCATCAATCCATGATCTATTCCAGGCATG 340
Db 264 ArgArgHisPro-----CysAlaProAlaAsp 272
QY 341 TCAGCCGCTGAGCTCTTACAGAGGCTGGAGAGG-----GGCAGGCCAGGT 394
Db 273 SerProArgSerGly-----GlyArgArgAspProGlySerProGly 287
QY 395 CA-----GAGTGGGTAGGGGCTCCAGACAGGCC----- 427
Db 288 ArgArgPheArgValSerAlaProLeuAlaValArgProProAlaAspGlyArgLeuPhe 307
QY 428 -----CTCCCTGTCTGGGCGCAATGCTCTGTGAGACTTGCC 469
Db 308 ArgAlaSerGlyPheAlaProCysArgGlyLeuLeuSerCysMetProAspSerAla 326

RESULT 5
US-09-252-991A-22317
; Sequence 22317, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22317
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22317

Alignment Scores:
Pred. No.: 0.00131 Length: 561
Score: 125.50 Matches: 54
Percent Similarity: 37.3% Conservative: 18
Best Local Similarity: 28.0% Mismatches: 55
Query Match: 13.6% Indels: 66
Gaps: 8

```





```
FILE REFERENCE: 407T-895820US
CURRENT APPLICATION NUMBER: US/09/477, 962
CURRENT FILING DATE: 2000-01-05
PRIOR APPLICATION NUMBER: 60/115, 435
PRIOR FILING DATE: 1999-01-06
PRIOR APPLICATION NUMBER: 60/118, 848
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 133
SOFTWARE: Patent In Ver. 3.0
SEQ ID NO: 107
LENGTH: 935
TYPE: PRT
ORGANISM: Streptomyces verticillius
FEATURE:
OTHER INFORMATION: ORF16
US-09-477-962-107

Alignment Scores:
Pred. No.: 0.00464 Length: 935
Score: 120.50 Matches: 56
Percent Similarity: 37.2% Conservative: 18
Best Local Similarity: 28.1% Mismatches: 59
Query Match: 13.1% Indels: 66
DB: 2 Gaps: 11

US-10-071-510A-16 (1-493) x US-09-477-962-107 (1-935)
QY 436 GACAGGAGGAGGCTGCTGCTGAGACCCCTTGACCTTGAGCGGCGCTGCGCCCT 377
DB 508 ACPARGALARGVALTHRALAGUALAProAlaAlaValProAlaAlaPro 527
QY 376 TCTCCACAGGCC-----CTGCTAGACGACTCAGCGCTCAGCA 338
DB 528 AAlaSerLarProAlaArgArgGluAlaGluLeuLeuAlaGlnValAlaArgValThr 547
QY 337 TGC-----CTGGGAATAGGATCATGCGATGATGAGAATCTGACTTTGGGATGTTG 284
DB 548 CyArgValLeuLylleGlyAlaVal-----GluProAspMetLeuLeu 563
QY 284 ----- 284
DB 564 ACPALAGLYAlaThrSerValGluLeuValArgLeuAlaThrAlaLeuGluGluLeu 583
QY 283 -----TTGGTGAAGAACCATTTTCTTCCCGGAATCCATCCCTCTGC 239
DB 584 GlyLeuAspThrAspLleGluGluLeuAlaAspPro--SerValAlaValIleVal 602
QY 238 AGATGCCAGTTTCTGATGAGGCTCAGCA-----GACTTTCTTCCAGCAGCCTC 191
DB 603 GlyArgHleLeuLylArgArg-ThrAlaProProAlaArgAspProLeuProProAlaSe 622
QY 190 A-----AGCCCGCCAGCAGCATCTGCCAGCTCCCTGAGCTCTCTGCGACACGCA 140
DB 622 ValAlaPheAlaPheArgLysValLeuPro-----AlaProProAlaProLylPr 639
QY 139 GCACCTCGCGAGAAAGACTTCTCCATCAGCAGCAGC-----CTGGCGCTTCCACAG 86
DB 639 oValProProAlaSerValProProAlaProAlaSerValProProAlaSerGluSerSe 659
QY 85 GACAGCTGGGCTCTTCTCCGGGA-----ATTCT 56
DB 659 rProLeuAlaProProAlaProLylProValProProthnProValProProAlaSerVa 679
QY 55 GCCACCAAGCTCTCAAACTCGGCTCTTGGGACTCGGCATCCCGCGGCGCGCG 1
DB 679 lProProAlaSerGlyAlaAlaPro-----HisValProProAlaPro 693

RESULT 11
US-08-818-112-143
Sequence 143, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
```

```
APPLICANT: Skeiky, Yasser A.W.
APPLICANT: Dillon, Devin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818, 112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121, 411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 143:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-818-112-143

Alignment Scores:
Pred. No.: 0.00305 Length: 174
Score: 120.00 Matches: 45
Percent Similarity: 31.1% Conservative: 2
Best Local Similarity: 29.8% Mismatches: 58
Query Match: 13.0% Indels: 46
DB: 2 Gaps: 8

US-10-071-510A-16 (1-493) x US-08-818-112-143 (1-174)
QY 411 CCCCTACCACTCTTGACTGGGCGCTG-----CCCTTCTCCACAGCCCTGC 361
DB 29 ProSerProProSerProProThrGlyThrValProArgAlaLeuLeuPro----- 45
QY 360 TAGACAGACTCAGCGGTGACGATGCTGGGAATGATGATGATGATGATGAAATC 301
DB 46 -----ProTrp----- 47
QY 300 CTGACTTGGAGATGTGTGTGTAAGAACCATTTTCTTCCCGGAATCCATTCATCTCT 241
DB 48 ---LeuAlaGlyThrProProAlaProProValProPrometAlaProLeuPro----- 64
QY 240 GCAGATGCCAGTTTCTGATGAGGCTCAGCAGACTTTCTTCCAGCAGCCTCAAGGCCGCC 181
DB 65 -----ProAlaAlaProLeuPro--- 70
QY 180 ACGACTTCCAGCTCCTGAGCTCTCTGACACGAGCAGCAGCCTCCGAGAAAGACT 121
DB 71 -----ProLeuProProLeuProProLeuProThrSerHisProProArgProPro 87
QY 120 TCTCCA-----TCACGACGAGCGCTGCG-----CTTCCACCAAGGAGA 82
DB 120 TCTCCA-----TCACGACGAGCGCTGCG-----CTTCCACCAAGGAGA 82
```

Db 88 AlaProProAlaProProAlaProAlaCysProPheValProValProProAlaPro 107  
QY 81 GCTGGGCTCTCTCTCG-----GGAATTCTGCCACGCTCT 43  
Db 108 ProLeuProProSerProProThrgluLeuProAlaAspAlaAlaCysProProAlaPro 127  
QY 42 CAACTCGGCTCTTGGGACTGGGCAATCCCGC 10  
Db 128 ProAlaProProLeuAlaProProSerProPro 138

RESULT 12  
US-08-818-111-138  
Sequence 138, Application US/08818111  
Patent No. 6338852  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonia  
APPLICANT: Houghton, Raymond  
APPLICANT: Vardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,111  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.4176  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 138:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 174 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-818-111-138

Alignment Scores:  
Pred. No.: 0.00305 Length: 174  
Score: 120.00 Matches: 45  
Percent Similarity: 31.1% Conservative: 2  
Best Local Similarity: 29.8% Mismatches: 46  
Query Match: 13.0% Indels: 58  
DB: 2 Gaps: 8

US-10-071-510a-16 (1-493) x US-08-818-111-138 (1-174)  
QY 411 CCCCTACCCACTTGTGACTGGGGCTCG-----CCCCTTCTCCACAGCCCTGC 361  
Db 29 ProSerProProSerProProThrglyTyrValProAlaAlaLeuLeuPro----- 45  
QY 360 TAGACAGACTCAGCGCTGCAGATGCTGGGAATAGATGATGAGATTGATGAGAAATC 301  
Db 46 -----ProTrr----- 47

QY 300 CTGACTTTGGAGTGTGTTGGTGAACATTTTCTTCCCGAATCCATTCCTCT 241  
Db 48 ---LeuAlaGlyThrProProAlaProProValProPrometAlaProLeuPro----- 64  
QY 240 GCAGATGCAGATTCTGATGAGCTCAGCAGACTTTCTTCCAGCAGCTCAAGCCCGCC 181  
Db 65 -----ProAlaAlaProLeuPro----- 70

QY 180 ACGACTTGCAGCTCTCTGAGCTCTCTGCAACGACGACGACCTCGGAGAAAGAT 121  
Db 71 -----ProLeuProProLeuProProLeuProThSerHisProProArgrProPro 87  
QY 120 TCTCCA-----TCACGAGCCAGCCCTGG-----CTTCACACGAGGACA 82  
Db 88 AlaProProAlaProProAlaProProAlaCysProPheValProValProProAlaPro 107  
QY 81 GCTGGGCTCTCTCTCG-----GGAATTCTGCCACGCTCT 43  
Db 108 ProLeuProProSerProProThrgluLeuProAlaAspAlaAlaCysProProAlaPro 127  
QY 42 CAACTCGGCTCTTGGGACTGGGCAATCCCGC 10  
Db 128 ProAlaProProLeuAlaProProSerProPro 138

RESULT 13  
US-09-056-556-143  
Sequence 143, Application US/09056556  
Patent No. 6350456  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
NUMBER OF SEQUENCES: 241  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 174 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-056-556-143

Alignment Scores:  
Pred. No.: 0.00305 Length: 174  
Score: 120.00 Matches: 45  
Percent Similarity: 31.1% Conservative: 2  
Best Local Similarity: 29.8% Mismatches: 46  
Query Match: 13.0% Indels: 58







**THIS PAGE LEFT BLANK**



Score:	156.50	Matches:	47
Percent Similarity:	38.4%	Conservative:	6
Best Local Similarity:	34.1%	Mismatches:	34
Query Match:	17.0%	Indels:	51
DB:	5	Gaps:	6

Score:	148.00	Matches:	43
Percent Similarity:	36.2%	Conservative:	8
Best Local Similarity:	30.5%	Mismatches:	42
Query Match:	16.0%	Indels:	48
DB:	4	Gaps:	5

US-10-071-510A-16 (1-493) x US-10-437-963-116465 (1-263)

QY	411	CCCCATCCCACTCTTGACCTGGGGCCCTTCTCCACAGCCCTGCTAGACAGAC	352
		:::	
Db	62	ProAlaProThrThrProProProAla---ProThrThrProProProAlaProThr---	79
QY	351	TCAGCGCGCAGCATGCTGGGAATGAGATCCATGGATGTGATGAAATCCTGACTTGG	292
Db	79	-----	79
QY	291	GGATGTGTGTGTGTAAGAAACATTTTCTTCCCGCAATCCATTCCTCTGACAGATGCC	232
		:::                :::	
Db	80	-----ThrProProProAlaProThr-----	87
QY	231	AGTTTGTGATGAGCTCAGACACTTCTTCCAGACGCTCAGAGCCGCGCAGACTGTG	172
Db	88	-----ProProProSerProProAlaThrProPro	97
QY	171	CCAGCTCCCTGAGCTCTCTCT-----GCACCGAGGCGACACCTCCGAGAAAGACTCT	118
Db	98	ProAlaProThrThrProProProSerProProSerGlnProProProAlaProAlaThr	117
QY	117	CCATCACCAGCAGCGCTCGCTTCCACCGAGGAGACGTGG-----CTTCCTTCTCGG	64
Db	118	ProProProSerProProProAlaThrProProProAlaProAlaThrProProProSerPro	137
QY	63	GGATTCCTGCCACGACCTCTCAAACTGGGCTCTTGGAGCTGGGATCCCCGGCCGG	4
		:::       :::	
Db	138	ProLeuAlaProProProAlaThrProProProProAlaThrProProProAla***Arg	157
QY	3	CCG 1	
Db	158	Pro 158	

RESULT 4  
US-10-425-115-263988  
Sequence 263988, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovacic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222) B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 263988  
LENGTH: 164  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(164)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_172372C.1.pep  
US-10-425-115-263988

Alignment Scores:  
Pred. No.: 0.00291 Length: 164  
Score: 144.50 Matches: 46  
Percent Similarity: 36.4% Conservative: 12  
Best Local Similarity: 30.5% Mismatches: 56  
Query Match: 15.7% Indels: 38

```

DB: 4 Gaps: 7
US-10-071-510A-16 (1-493) x US-10-425-115--263988 (1-164)
QY 425 GCCTGTGCTGAGAGACCCCTACCCACTCTTGACTCGGGGGCTGGCCCCCTTTCCTCCACAGC 366
Db 16 SercyslaiaaIagInGInserProProaIaGInaIaProProThrProaSnlaaProaIa 35
QY 365 CCTGCTAAGACACTCAGCGCTGACGATGCTT---GGGAATAG----- 325
Db 36 ProProaIaAaSnSerProProGInaIaProProaIaGInaIaAaSnProProProaIaProGIn 55
QY 324 GATTCATGGGATTTGATVAGAAATCTGACTTTGGGATGTTTGTGTGAAAAACCATTTTCT 265
Db 56 AlaPro-----ProaIaGly 60
QY 264 TCCCGGAATCCACTTCATCCTCTGACAGATGCCAGTTCTGATGAGGCTCAGCAGACTTT 205
Db 61 AaSnProProProIaIaProThrAspThrProSer----- 71
QY 204 CTTCCAGCAGCCTCAAGGCCCGCCACGACTCTGCCAGCTTCCTGAGCTCT---CCTGCA 148
Db 72 ---ProaIa-----ProThrThrProProProaIaProThrThrProSerProaIa 87
QY 147 CCACGGCAGCACCCTCCGGAGAAGACTTCCATCAGCAGCAGCCCTCGCTTCACCA 88
Db 88 ProThrThrProProProaIaProThrThrProProProSerProProaIaSerProPro 107
QY 87 GGCAGAGCTGG-----CCTCCTTCTCCGGGAATTTCTGCCACAGCCTCTCAAACTGG 34
Db 108 ProaIaProaIaThrProProProSerProProMetAlaProPro**AlaThrProPro 127
QY 33 CCTCTTGGGACTCGGCATCCCCGGCGCGCG 1
Db 128 ProProaIaThrProProProProaIaIaPro 138

RESULT 5
US-10-437-963-168762
/ Sequence 168762, Application US/10437963
/ Publication NO. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Zhou, Yinhua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 168762
/ LENGTH: 437
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(437)
/ OTHER INFORMATION: unsure at all xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_67245C.1.pcp
US-10-437-963-168762

Alignment Scores:
Pred. No.: 0.00342
Score: 144.50
Percent Similarity: 32.7%
Best Local Similarity: 27.6%
Query Match: 15.7%
Length: 437
Matches: 54
Conservative: 10
Mismatches: 43
Indels: 89

```

[illegible]

```

US-10-437-963-189874

Alignment Scores:
Pred. No.: 0.0038      Length: 148
Score: 143.00         Matches: 36
Percent Similarity: 41.2%      Conservative: 6
Best Local Similarity: 35.3%    Mismatches: 38
Query Match: 15.5%           Indels: 22
DB: 4                      Gaps: 4

US-10-071-510A-16 (1-493) x US-10-437-963-189874 (1-148)

OY 282 TGGTGAACAACATTTTCTTCCCGGAATCCACTTCATCTCTGAGATGCCAGTTTCGA 223
    ||| :||| ||||| :||| ||| :|||
DB 45 TTPAAGAGProSerSerSerSerSerProThProAlaPro**Pro----- 60

OY 222 TGAGGCTCAGCAGACTTCTTCCAGCAGCCTCAAGGCCCGCCACGACTCTCCAGCTCCC 163
    ||| :||| ||||| :||| ||| :|||
DB 61 -----ProAlaPro**ProProSerProAlaProProPro 73

OY 162 TGAGTCTCTCTGACACGAGCAGCAGCCTTCGCGAAGAAGACTTCTCCA----- 115
    ||||| :||| ||||| :||| ||| :|||
DB 74 ATgThProProSerProSerProProProProProProProProProProPro 93

OY 114 -----TCACACGACGACCTCGGCTTCACACGAGGACAGCTGGGCTCTTCT 67
    ||| :||| ||||| :||| ||||| :|||
DB 94 ThrgIyaP**ProProProSerProAlaProProAlaThr-----ProProSer 111

OY 66 CCGGGAATTCCTGCCACGACGACCTCTCAAACTCGGCTCTTGAGACTCGGATCCCGCGGC 7
    ||| :||| ||||| :||| ||| :|||
DB 112 ProSerProAlaProProProProProSerProAlaPro---ProProSerProProGly 130

OY 6 CGGCGC 1
    |||||
DB 131 ATgPro 132

RESULT 7
US-10-437-963-161137
; Sequence 161137, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204366
; SEQ ID NO 161137
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(501)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_60349C.1.pep
US-10-437-963-161137

Alignment Scores:
Pred. No.: 0.00561      Length: 501
Score: 142.00         Matches: 51
Percent Similarity: 35.7%      Conservative: 9
Best Local Similarity: 30.4%    Mismatches: 54
Query Match: 15.4%           Indels: 54

```

```

DB: 4 Gaps: 6
US-10-071-510A-16 (1-493) x US-10-437-963-161137 (1-501)
QY 476 AGAGCCCTGGCCCAAGTCTCAGAGCAGAGATGTTGGCCCAAGAGGAGGAGGCGCTGTGCT 417
   |||||:|||||
Db 108 AtgThrPheProProProProSer-----Pro 116
QY 416 GGAGACCCCTTACCACCTCTTGACCTGGGAGCCCTGGCCCTTCTCCAGAGCCCTGTAGA 357
   |||||:|||||
Db 117 ProProProAlaProProProProProProProProProProProProProProProPro 136
QY 356 CAGACTCAGCGCTGACGATGCTGGGAAATAGATCCATGGAGATTGATGAAGAAATCCTGA 297
   |||||:|||||
Db 137 ThrThr**ProProThrSerProProPro--ProProThr----- 149
QY 296 CTTTGGGATGTTGTGTGTAAGAAACATTTTCTTCCCGAATCCATTCCTCTGACG 237
   |||||:|||||
Db 150 -----ProProThrGlnProSerAlaProLeuProAlaPro---- 161
QY 236 ATGCCAGTTTGTATGAGGCTCAGACACTTCTTCCAGCAGCCTGAAGCCCGCCAGCA 177
   |||||:|||||
Db 162 -----ProSerSerSerProProProProProProThrt 170
QY 176 CTCTGCGACCTCCTTAGCTCTCTGACACGAGCAGACACCTCCGAGAAACTTCTC 117
   |||||:|||||
Db 170 hrProProProProSerThrProPro-----ProProProProThrSerP 185
QY 116 CATCACCAGCAGACCTCGCGCTCCACAGGAGCAGCTGGGCGCTCTCTC----- 67
   |||||:|||||
Db 185 rSerThrProAlaPro**ProProProSerSerSerProProThrtThrProProT 205
QY 66 -----CCGGGAATTCGTGCGACAGCCTCTCAACTCGGCTCTTGAGACTC 21
   |||||:|||||
Db 205 hrSerAlaProProProProProProThrProProSerSerSerProPro-----Thrt 223
QY 20 GGCATCCCGCGGCGCGCG 1
   |||||:|||||
Db 223 erProProProSerThrPro 229

RESULT 8
US-10-425-115-289952
; Sequence 289952, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 289952
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; LOCATION: (1)..(224)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: MRT4577_27523C.1.pdp
US-10-425-115-289952

Alignment Scores:
Pred. No.: 0.00593 Length: 224
Score: 141.00 Matches: 43
Percent Similarity: 34.1% Conservative: 53
Best Local Similarity: 31.2% Mismatches: 53
Query Match: 15.3% Indels: 38

```

```

DB: 4 Gaps: 5
US-10-071-510A-16 (1-493) x US-10-425-115-289952 (1-224)
QY 414 AGACCCCTTACCCACCTTGTGACCTGGGGCTGGCCCTTCTTCCACAGCCCTGTAGACA 355
   |||||:|||||
Db 38 ArgProLeuProLeuProAlaPro--AlaProProPro**ProProProAla----- 54
QY 354 GACTCAGCGCTGACGATGCTTGGGAATAGATTCATGGGATTTGATGAAGAAATCCTGACT 295
   |||||:|||||
Db 55 -----Pro----- 55
QY 294 TTGGGATGTTGTGTAAGAAACATTTTCTTCCCGAATCCATCTCATCTCTGACAT 235
   |||||:|||||
Db 56 -----ProPro**ProProProGluProLeuProAlaProAla---- 68
QY 234 GCCAGTTTGTATGAGGCTCAGACACTTCTTCCAGCAGCCTCAAGCCCGCCAGCACT 175
   |||||:|||||
Db 69 -----ProProProTyProProProAlaProProProAla----- 80
QY 174 CTGCCAGCTCCCTGAGCTCTCTTGTGACCAAGCAGACACCTTCCGAGAAACTTCTCA 115
   |||||:|||||
Db 81 ---ProProPro**ProProProAlaProProProTyProProProLeuProProPro 99
QY 114 TCACAGCCAGCGCTGCGCTTCCAGCAGGAGACAGCTGGGCGCTCTCTCCGGAATTTCTG 55
   |||||:|||||
Db 100 AlaProAlaProProPro**ProProProAlaProAlaProProProProProAlaArgPro 119
QY 54 CCACAGCCTCTCAAACTCGGCTCTTGGGACTCGGCAATCCCGGAGCGCGCG 1
   |||||:|||||
Db 120 ProProProProSerProProProProAlaProProGluProProProAlaArgPro 137

RESULT 9
US-10-425-115-263984
; Sequence 263984, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 263984
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; LOCATION: (1)..(231)
; OTHER INFORMATION: Clone ID: MRT4577_172369C.1.pdp
US-10-425-115-263984

Alignment Scores:
Pred. No.: 0.00596 Length: 231
Score: 141.00 Matches: 39
Percent Similarity: 35.3% Conservative: 10
Best Local Similarity: 28.1% Mismatches: 46
Query Match: 15.3% Indels: 44
DB: 4 Gaps: 4

US-10-071-510A-16 (1-493) x US-10-425-115-263984 (1-221)
QY 411 CCCTTACCCGACCTTGTGACCTGGGAGCTGGCCCTTCTCCACAGCCCTGTAGACAGAC 352
   |||||:|||||
Db 62 ProAlaProThrAlaThrProProProAlaProAlaProThrProProProAlaProThr--- 80
QY 351 TCACGCGCTGACGATGCTTGGGAATAGATTCATGGGATTTGATGAAGAAATCTGACTTTG 292
   |||||:|||||
Db 80 ----- 80

```

QY 291 GGATGTTGTTGGAACCATTTTCTCCCGAATCCATTCCATCTCTGAGATGCG 232  
DB 81 -----ThrProProAlaProThrProProPro 91  
QY 231 AGTTCTGATGAGGCTCAGCAGACTTTCTTCCAGAGCCTCAAGGCCCGCAGACTG 172  
DB 92 Ala-----ProThrThrProProProAlaProThrThr 102  
QY 171 CCAGCTCCCTGAGCTCTCTGACACGAGGACCCCTCCGAGAAAGCTTCCATCA 112  
DB 103 ProProProSerProProAlaSerPro-----ProProAlaProThrThrProPro 119  
QY 111 CCAGCAGAGCCCTGCGCTTCCACCAGGAGACGCTGG-----CCTTCCTTCTCCGGGAAT 58  
DB 120 ProSerProProAlaSerProProProAlaProAlaThrProProProSerProProMet 139  
QY 57 CTGCCACGAGCCTCTCAAACTCGGCTCTTGGGACTCGGCATCCCGGCGCGCG 1  
DB 140 AlaProProProAlaThrProProProProAlaThrProProProProAlaAlaPro 158  
RESULT 10  
US-10-739-930-10816  
; Sequence 10816, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739, 930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 10816  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(238)  
; OTHER INFORMATION: unsure at all Xaa locations  
; OTHER INFORMATION: Clone ID: TRIAE-23ABR03-C730\_1224.p  
US-10-739-930-10816  
Alignment Scores:  
Pred. No.: 0.00599 Length: 238  
Score: 141.00 Matches: 39  
Percent Similarity: 42.6% Conservative: 7  
Best Local Similarity: 36.1% Mismatches: 40  
Query Match: 15.3% Indels: 22  
DB: 5 Gaps: 4  
US-10-071-510A-16 (1-493) x US-10-739-930-10816 (1-238)  
QY 273 CCATTTTCTCCCGAATCCATTCCTCGAGGATCCAGTTTGTGATGAGGCTCA 214  
DB 56 ProAlaGlyAsnProProProAlaProSerAlaProProAlaProThrThrProProGln 75  
QY 213 GCAGACTTCTTCCAGCAGCC-----TCAAGGCCCGCCCAAGACTTGCCA 169  
DB 76 AlaProGlnAlaProAlaThrProProProAlaProGlnAlaProAlaThrProProPro 95  
QY 168 GCTCCCTGAGCTCTCTGCA----- 148  
DB 96 AlaPro\*\*\*ThrProProAlaThrProProProAlaProThrThrProProProAlaPro 115  
QY 147 CCAGCGAGCAGCAGCCCTCCGAGAAAGACTTCCATCAACGAGCAGCCCTGCGTCCACCA 88  
DB 116 ProThrGlnProProProAlaProThrThrThrThrThrThrThrThrThrThrThrThr 135  
QY 87 GGGACAGCTGG-----CCTCCTTCTCCGGGAATTCCTCCACGAGCCTCAAACTCGG 34  
DB 136 ProAlaProAlaThrProProProSerProProMetSerProProProAlaThrProPro 155

QY 33 CCTTTGGGAGCTGGGATCCCCCG 10  
DB 156 PrometAlaThr-----ProPro 161  
RESULT 11  
US-10-425-115-263989  
; Sequence 263989, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425, 115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 263989  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(239)  
; OTHER INFORMATION: unsure at all Xaa locations  
; OTHER INFORMATION: Clone ID: MRT4577\_172373C.1.dep  
US-10-425-115-263989  
Alignment Scores:  
Pred. No.: 0.00599 Length: 239  
Score: 141.00 Matches: 39  
Percent Similarity: 35.3% Conservative: 10  
Best Local Similarity: 28.1% Mismatches: 46  
Query Match: 15.3% Indels: 44  
DB: 4 Gaps: 4

US-10-071-510A-16 (1-493) x US-10-425-115-263989 (1-239)  
QY 411 CCCCTACCCACTCTTGACCTGGGCGCTTCTTCCACAGCCCTGTAGACAGC 352  
DB 69 ProAlaProThrAlaThrProProProAlaProThrThrProProProAlaProThr--- 87  
QY 351 TCACGCGTCAGAGATGCTGGGAATGATCCATGAGATGAGAAATCTGACTTG 292  
DB 87 ----- 87  
QY 291 GGATGTTGTTGGAACCATTTTCTCCCGAATCCATTCCTCTGCAATGCC 232  
DB 88 -----ThrProProProAlaProThrThrProProPro 98  
QY 231 AGTTCTGATGAGGCTCAGCAGACTTCTTCCAGCAGCCTCAAGGCCCGCAGACTG 172  
DB 99 Ala-----ProThrThrProProProAlaProThrThr 109  
QY 171 CCAGCTCCCTGAGCTCTCTGACACGAGGACCCCTCCGAGAAAGACTTTCATCA 112  
DB 110 ProProProSerProProAlaSerPro-----ProProAlaProThrThrProPro 126  
QY 111 CCAGCAGCCCTGCGCTTCCACAGGAGCAGCTGG-----CCTCCTTCTCCGGGAAT 58  
DB 127 ProSerProProAlaSerProProProAlaProAlaThrProProProSerProProMet 146  
QY 57 CTGCCACGAGCCTCTCAAACTCGGCTCTTGGGACTGGGATCCCGGCGCGCGCG 1  
DB 147 AlaProProAlaThrProProProProAlaThrProProProProAlaAlaPro 165  
RESULT 12  
US-10-767-701-40674  
; Sequence 40674, Application US/10767701



```

; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 40674
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C80510_1.pep
; US-10-767-701-40674

Alignment Scores:
Pred. No.: 0.00982 Length: 155
Score: 138.00 Matches: 36
Percent Similarity: 38.5% Conservative: 48
Best Local Similarity: 34.6% Mismatches: 4
Query Match: 15.0% Indels: 16
DB: 4 Gaps: 2

US-10-071-510A-16 (1-493) x US-10-767-701-40674 (1-155)
QY 282 TGGTGAACCATTTTCTTCCCGAATCCACTTCTGCAAGTCCGATTTCTGA 223
DB 7 TTPProArgProLeuSerSerProArgAlaProSerProProAla----- 23
QY 222 TGAAGCTGACGACACTTTCTTCCAGACAGCTCAAGCCCGCCAGACTCGACGCTCC 163
DB 24 -----AlaProProAlaProProProProProProAlaProProProPro 40
QY 162 TGAGCTCTCTCTGACACGCG-----CAGCACTT 133
DB 41 LeuAlaProProAlaProArgGlyAlaArgArgProArgGlyArgProProAlaPro 60
QY 132 CCGGAGAAAGCTTCTTCATCAGCAGCCGCTGCGCTTCCAGCAGGAGACGCTGGCT 73
DB 61 ProAlaGlyArgAlaAlaAlaProAlaProAlaArgProProAlaGlyArg 80
QY 72 CCTTCCCGGAGATTCTGCGACGACGCTCAACTGCGCTTGGAGCTGGGCAATCCC 13
DB 81 ProGlyProGlyAlaAlaProGlyProProProSerArgAlaAlaProThrArgArgPro 100
QY 12 CCGGCGCGGCG 1
DB 101 ArgAlaProPro 104

RESULT 13
US-10-767-701-36014
; Sequence 36014, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 36014
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure

```

```

; LOCATION: (1)..(189)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C79158_1.pep
; US-10-767-701-36014

Alignment Scores:
Pred. No.: 0.0101 Length: 189
Score: 138.00 Matches: 37
Percent Similarity: 42.3% Conservative: 4
Best Local Similarity: 38.1% Mismatches: 34
Query Match: 15.0% Indels: 22
DB: 4 Gaps: 4

US-10-071-510A-16 (1-493) x US-10-767-701-36014 (1-189)
QY 267 TCTTCCCGAATCCATTCATCTGCGAGATGCGAGTTTGTATGAGCTCAGCAGAC 208
DB 3 AlaSerProSerProLeu----- 8
QY 207 TTTTTCAGACGCTCA-----AGCCCGCCAGACTTGCAGCTCCCTGA 160
DB 9 LeuProProProAlaSerArgArgProProArgProValProProProProPro 28
QY 159 GCTCTCTGCAACCGGACGACACCTCCGGAGAGACTTTCATCAGCA-----GCC 106
DB 29 ProProProSerProArgSerProAlaProSerPro**ProProSerProProArgPro 48
QY 105 AGCCTGCGCTTCCAGCAGGAGACGCTGCGCTTCTCCG-----GGAATTCGCA 52
DB 49 ProProAlaSerProProProAlaSerGlyProProSerProValLeuGlyArgArgPro 68
QY 51 CCAGCCTCTCAACTCGGCTTCTGGAGCTCGGCAATCCCCGGGCGCGCC 1
DB 69 ProValArgAlaCysArgProCysArgGlyProThrAlaProAlaProArgPro 85

RESULT 14
US-10-369-493-3962
; Sequence 3962, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3962
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Neurospora crassa
; US-10-369-493-3962

Alignment Scores:
Pred. No.: 0.0124 Length: 639
Score: 138.00 Matches: 39
Percent Similarity: 43.8% Conservative: 3
Best Local Similarity: 40.6% Mismatches: 36
Query Match: 15.0% Indels: 18
DB: 4 Gaps: 4

US-10-071-510A-16 (1-493) x US-10-369-493-3962 (1-639)
QY 255 CCACTTCATCTCTGAGATGCGAGTTTCTGATGAGGCTCAGCAGACTTCTT---CCA 199
DB 423 ProLeuProProGlyAlaPro-----GlyProAlaProProProLeuProPro 437

```

```
QY 198 GCAGCCTCAAGCGCCGCAAGACTTGCACGCTCCTGAGCTCTCTGACCAAGCGGCG 139
DB 438 AAlaSerSerArgProProProMetLeuProThr-----ArgSerProAlaProProGln 455
QY 138 CAGCCCTCGGAGAGAACTTCTCCATCAGCAGCGCCCTGCGGCTTCA-----91
DB 456 AAlaProProLeuProThrSerAsnAlaProProProProProAlaThrGlnAla 475
QY 90 -----CCAGGAGACGCTGGGCTCTCTTCCGGGAATTCTGACCA 49
DB 476 ProProProProProLeuProProAlaThrSerAlaProProProProProProAlaProPro 495
QY 48 GCCTCTCAAACTCGGCTCTTGGGACTCGGCATCCCGGCGCGCG 1
DB 496 AAlaProProAlaProProProProAlaAlaAlaAlaProProProPro 511

RESULT 15
US-10-425-115-229737
; Sequence 229737, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 229737
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(180)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_141114C.1.pep
US-10-425-115-229737

Alignment Scores:
Pred. No.: 0.011 Length: 180
Score: 137.50 Matches: 38
Percent Similarity: 31.3% Conservative: 3
Best Local Similarity: 29.0% Mismatches: 39
Query Match: 14.9% Indels: 51
DB: 4 Gaps: 5

US-10-071-510A-16 (1-493) x US-10-425-115-229737 (1-180)
QY 393 CCTGGGCGCTGGCCCTTCTCCAGCGCCGTGACAGACTCAAGCGGTGACGATGCC 334
DB 1 ProAlaHisArgProLeuLeuProProProCys-----11
QY 333 TGGGAATAGATCAATGGATTGATGAGAAATCTGACTTTGGGATGTTGTTGTAATA 274
DB 11 -----11
QY 273 CCATTTTCTCCCGAATCCACTTCATCTCTGACAGATGCCAGTTTCTGATGAGGCTCA 214
DB 12 -----HisSerProLeuProPro-----18
QY 213 GGAAGACTTCTTCCAGCAGCTCAAGCGCCGACAGACTCTGCAAGCTCCGAGCTCCT 154
DB 19 -----Pro**ProArgArgProProGlnPro**Pro**Pro-----Pro 32
QY 153 CCTGACACAGCGAGCAGCCCTCGGAGAGACTTCTCATACAGCAGCCCTGCGGCTT 94
DB 33 ProCysProProProGlnProPro-----ArgAspSerProThrProProProProThrPro 51
```

```
QY 93 CCACCAAGGAGACAGCTGGGCTCTCTTCTCCGGAATTCTGCCACAGCCTCTCAAACTCGG 34
DB 52 ProProProProProHisProProProProProProHisProProProSerProProAsn 71
QY 33 CCTTTGGGACTCGGCATCCCGGAGCGCGCG 1
DB 72 Pro-----ProProHisProProProProPro 80
```

Search completed: April 29, 2006, 04:09:01  
Job time : 124 secs



[illegible]

```

RESULT 2
US-11-036-256-89
; Sequence 89, Application US/11036256
; Publication No. US20060026719A1
; GENERAL INFORMATION:
; APPLICANT: KIELISZEWSKI, MARCIA
; APPLICANT: XU, JIANFENG
; TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND
; FILE REFERENCE: 27211/04130
; CURRENT APPLICATION NUMBER: US/11/036,256
; PRIOR FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: 60/502,562
; PRIOR FILING DATE: 2004-08-18
; PRIOR APPLICATION NUMBER: 60/582,027
; PRIOR FILING DATE: 2004-06-22
; PRIOR APPLICATION NUMBER: 60/536,486
; PRIOR FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 89
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-11-036-256-89

Alignment Scores:
Pred. No.:      0.00271      Length:      411
Score:          128.50      Matches:      45
Percent Similarity:      35.0%      Conserved:      5
Pest Local Similarity:    31.5%      Mismatches:    52
Query Match:      13.9%      Indels:      41
DB:              7          Gaps:          7

US-10-071-510A-16 (1-493) x US-11-036-256-89 (1-411)

QY      411 CCCCTACCACCTCTTGACCTGGGGGCTGG-----CCCTCTCCACAGCCC 364
      ||| ||||| ||||| ||||| ||||| ||||| |||||
Db      94 ProThrProThrProProProGlyProHisSerProProProProLeuSerProSerPro 113
      ||| ||||| ||||| ||||| ||||| ||||| |||||
QY      363 TGTCTAGACAGACTCAGCGCTGCAGATGCTGGGATAGATGATCGATGATGAGAA 304
      ||| ||||| ||||| ||||| ||||| ||||| |||||
Db      114 ThrProThrPro----- 117

```

```

Oy      303 ATCTGACTTTGGAGATGTTGGTGGTAAACCAATTTTCTCCCGAATCCACTTCCATCC 244
          |||||
Db      118 -----ProIeuGlyProHiseSerProProHthrLeuSerProSerProHthr 135
          |||||
Oy      243 TCTGCAGATGCSAGTTCTGATGAGAGCTCAGAGACTTTCTTCCAGACGCTCAAGGCC 184
          |||||
Db      136 ProPro-----ProGlyProHiseSerPro 143
          |||||
Oy      183 GCCACAGACTGTG-----CSAGCTCCCTGAGCTCTCTGACACACAGGACAGACCT 133
          |||||
Db      144 ProProProIeuSerProSerProHthrPro---ThrProIeuGlyProHiseSerPro 162
          |||||
Oy      132 CCGGAGAAAGACTTTTCATCACCACGACCCCTGCGCTTCCACAGGAGACAGTGGGACT 73
          |||||
Db      163 ProProHthrIeuSerProSerPro---ThrProHthrProProProGlyValProGlyPro 181
          |||||
Oy      72 -----CCTTTCGCCGGGAATTCTGCACACAGCTCTCAAACTCGGCTCTTTGGGACTCG 19
          |||||
Db      182 HiseSerProProProIeuSerProSerProHthrProHthrProIeuGlyProHise 201
          |||||
Oy      18 CATCCCCG 10
          |||||
Db      202 SerProPro 204

```

```

RESULT 3
US-11-232-440-19
; Sequence 19, Application US/11232440
; Publication No. US20060068434A1
; GENERAL INFORMATION:
; APPLICANT: STORER, JAY
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING CANCER USING
; FILE OF INVENTION: COMPONENTS OF THE U2 SPLICEOSOMAL PARTICLE
; FILE REFERENCE: MTP-031
; CURRENT APPLICATION NUMBER: US/11/232,440
; CURRENT FILING DATE: 2005-09-21
; PRIOR APPLICATION NUMBER: 60/612,310
; PRIOR FILING DATE: 2004-09-21
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 19
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-232-440-19

Alignment Scores:
Pred. No.: 0.00404 Length: 464
Score: 126.50 Matches: 52
Percent Similarity: 38.8% Conservative: 12
Best Local Similarity: 31.5% Mismatches: 66
Query Match: 13.7% Indels: 35
DB: 7 Gaps: 9

US-10-071-510A-16 (1-493) x US-11-232-440-19 (1-464)
Oy 440 CCAGACAGGAGGAGGCGCTGCTGAGAGACCCCTTACCCACTTGTGACCTGGGCGCTGCG 381
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ProProthrdGlyProAlaProSerdlYProProGlyProProGlnInLeuProProAla 284
Oy 380 CCCCTTCCACAGCCCTCTGACAGACCTACCGCGCTGACGATGCTCGGAGATAGCATC 321
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 ProGlyValHisProProAlaProValValHisPro-----ProAlaSerGlyVal 301
Oy 320 CAT-----GGGATGTATAGAAATCCGTGACTTTGGATGTGTTGGGAAAC 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 HisProProAlaProGlyValHisProProAla-----ProGlyValHis 316
Oy 272 CATTTTCTTCCCGAATCCAC-----TTCCATCTCTGACAGATGCCAGTT 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 ProProAlaProGlyValHisProProthrSerGlyValHisProProAlaProGlyVal 336
227 TCTGATGAGGCTCAGCAGACTTTCTTCCAGCAGCCTCAGGCGCCGCCACGACTTGGCAG 166

```

```
Db 337 HisProProAlaProGlyValHis--ProProAla-----ProGlyValHisProProAla 354
Qy 167 CTCCTGAGCT---CCTCTGACACGAGCAGACCTT-----C 132
Db 354 lAProGlyValHisProProAlaProGlyValHisProProSerAlaGlyValHisP 374
Qy 131 CGAGAGAAGATTCTTCATGACACGAGCAGCTGCGCTTCCACGAGGAGACGCTGGG- 76
Db 374 roGlnAlaProGlyValHisProAlaAlaProAlaValHisProGlnAlaProGlyValH 394
Qy 75 --CTCTCTTCTCCGGGAATTCCTCCACGAGCCTCTAAACTCGGCTCTTGGAGCTGGG 18
Db 394 lAProProAlaProGlyValHisProGlnAlaProGlyValHisPro-----G 410
Qy 17 ATCCCCGGGGC 7
Db 410 lAProProGly 413

RESULT 4
US-11-232-440-67
; Sequence 67, Application US/11232440
; Publication No. US20060068434A1
; GENERAL INFORMATION:
; APPLICANT: STOECKER, JAY
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING CANCER USING
; FILE REFERENCE: MTP-031
; CURRENT APPLICATION NUMBER: US/11/232,440
; PRIOR FILING DATE: 2005-09-21
; PRIOR APPLICATION NUMBER: 60/612,310
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 67
; LENGTH: 464
; TYPE: PRF
; ORGANISM: Homo sapiens
US-11-232-440-67
```

```
Alignment Scores:
Pred. No.: 0.00404 Length: 464
Score: 126.50 Matches: 52
Percent Similarity: 38.8% Conservative: 12
Best Local Similarity: 31.5% Mismatches: 66
Query Match: 13.7% Indels: 35
DB: 7 Gaps: 9
```

US-10-071-510a-16 (1-493) x US-11-232-440-67 (1-464)

```
Qy 440 CCCAGACAGGAGGGGCTGTGGTGGAGAGCCCTTACCCACTTGTGACCTGGGGCTTGGC 381
Db 265 ProProThriGlyProAlaProSerGlyProProGlyProProGlnLeuProProProAla 284
Qy 380 CCTTCTCCACAGCCTGTGTAGACAGACTCAGCAGCTGAGATGCTGCGATAGATGATC 321
Db 285 ProGlyValHisProProAlaProAlaValHisPro-----ProAlaSerGlyVal 301
Qy 320 CAT-----GGGATTGATGAGAAATCTGACCTTGGGATGTTGTTGGTGAAGAAC 273
Db 302 HisProProAlaProGlyValHisProProAla-----ProGlyValHis 316
Qy 272 CATTTTCTCCGGAATCCAC-----TTCCATCTCTGACAGATGCCAGTT 228
Db 317 ProProAlaProGlyValHisProProThriSerGlyValHisProProAlaProGlyVal 336
Qy 227 TCTGATGAGCTCAGCACTTTCTTCCAGCAGCCTTGAAGCCGCCAGCACTGCTCCAG 168
Db 337 HisProProAlaProGlyValHis--ProProAla-----ProGlyValHisProProAla 354
Qy 167 CTCCTGAGCT---CTTCTGACACGAGGAGCAGCTT-----C 132
Db 354 lAProGlyValHisProProAlaProGlyValHisProProProSerAlaGlyValHisP 374
```

```
Qy 131 CGAGAGAAGATTCTTCATGACACGAGCAGCTGCGCTTCCACGAGGAGACGCTGGG- 76
Db 374 roGlnAlaProGlyValHisProAlaAlaProAlaValHisProGlnAlaProGlyValH 394
Qy 75 --CTCTCTTCTCCGGGAATTCCTCCACGAGCCTCTAAACTCGGCTCTTGGAGCTGGG 18
Db 394 lAProProAlaProGlyValHisProGlnAlaProGlyValHisPro-----G 410
Qy 17 ATCCCCGGGGC 7
Db 410 lAProProGly 413
```

```
RESULT 5
US-11-087-099-7887
; Sequence 7887, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; PRIOR FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 7887
; LENGTH: 647
; TYPE: PRF
; ORGANISM: Brassica napus
US-11-087-099-7887
```

```
Alignment Scores:
Pred. No.: 0.00613 Length: 647
Score: 124.50 Matches: 36
Percent Similarity: 48.4% Conservative: 10
Best Local Similarity: 37.9% Mismatches: 40
Query Match: 13.5% Indels: 9
DB: 7 Gaps: 6
```

US-10-071-510a-16 (1-493) x US-11-087-099-7887 (1-647)

```
Qy 273 CCATTTCTTCCCGAATCCA-----CTTCAATCTTGCAGATGCCAGTTTCTGATGA 220
Db 35 ProSerSerProProProProSerThriLeProThriSerProProSerSer----- 52
Qy 219 GAGTCAGCAGACTTTTCTCCAGCAGCTCAAGCCCGCCAGCACTGCTCCAGTCCCTGA 160
Db 53 ArgSerThriProSerAlaProProProSerProProThriProSerThriProGlySer-- 71
Qy 159 GCTCTCTCTCCACAGGAGCAGCAGCCTTCCGGAGAGACTTCTCA---TCACGAGCAGC 103
Db 72 ProProProLeuProGlnProSerProProAlaProThriProGlySerProProAla 91
Qy 102 CTGCGCTTCCACAGGAGCAGCTGGGCTTCTTCT---CCGGGAATTTGCCACAGCC 46
Db 92 ProValThriProProThriArgAnProProProSerAlaProGly-----ProProSer 109
Qy 45 TCTCAACTCGGCTTGGGAGCTCGGCAATCCCGGCGCGCGCC 1
Db 110 AsnProSerArgGlnGlyGlySerProArgProProSerSerPro 124
```

```
RESULT 6
US-11-188-298-18316
; Sequence 18316, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; PRIOR FILING DATE: 2005-07-22
; CURRENT APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
```

```
; SEQ ID NO 18316
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Brassica napus
US-11-188-298-18316

Alignment Scores:
Pred. No.: 0.00613      Length: 647
Score: 124.50          Matches: 36
Percent Similarity: 48.4%      Conservative: 10
Best Local Similarity: 37.9%    Mismatches: 40
Query Match: 13.5%           Indels: 9
DB: 7                     Gaps: 6

US-10-071-510A-16 (1-493) x US-11-188-298-18316 (1-647)
QY 273 CCATTTTCTTCCCGAATCCA-----CTTCCATCTCTGACAGATCCAGTTTCTGATGA 220
DB 35 ProSerSerProProProProSerThrIleProThrIleProProProSer----- 52
QY 219 GGCTGACAGACTTCTTCCAGCAGCCTCAAGCCCGCCACACACTCTGCCAGCTCCCTGA 160
DB 53 ArgSerThrProSerAlaProProProSerProThrProSerThrProGlySer--- 71
QY 159 GCTCCTCCGACACGACGACGACACCTCCGAGAAAGACTTCTCA---TCACACAGCAGC 103
DB 72 ProProProLeuProGlnProSerProProAlaProThrThrProGlySerProProAla 91
QY 102 CCTGCGCTTCCACAGGACGACAGCTGGGCTCTCT---CCGGAATTCGCCACAGCC 46
DB 92 ProValThrProProThrArgAnProProProSerValProGly-----ProProSer 109
QY 45 TCTCAACTCGGCGCTTGGGACTCGGCAATCCCGGCGCGCGCG 1
DB 110 AsnProSerArgGluGlyGlySerProArgProProSerSerPro 124

RESULT 7
US-11-036-256-21
; Sequence 21, Application US/11036256
; Publication No. US20060026719A1
; GENERAL INFORMATION:
; APPLICANT: KIELISZEWSKI, MARCIA
; APPLICANT: XU, JIANFENG
; TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND
; FILE REFERENCE: 27211/04130
; CURRENT APPLICATION NUMBER: US/11/036,256
; PRIOR FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: 60/602,562
; PRIOR FILING DATE: 2004-08-18
; PRIOR APPLICATION NUMBER: 60/582,027
; PRIOR FILING DATE: 2004-06-22
; PRIOR APPLICATION NUMBER: 60/536,486
; PRIOR FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 21
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid construct
US-11-036-256-21

Alignment Scores:
Pred. No.: 0.00742      Length: 183
Score: 123.00          Matches: 44
Percent Similarity: 34.8%      Conservative: 5
Best Local Similarity: 31.2%    Mismatches: 52
Query Match: 13.3%           Indels: 40
DB: 7                     Gaps: 7
```

```
US-10-071-510A-16 (1-493) x US-11-036-256-21 (1-183)
QY 411 CCCCTACCCACTTCTTGACCTGGGGCCCTG-----CCCTTCTCCACAGCC 364
DB 56 ProThrProThrProProProGlyProHisSerProProProLeuSerProSerPro 75
QY 363 TGCTGACAGACTCAGCGGTCGACGATCCTGGAAATGATCCATGGAGATTGATGAGA 304
DB 76 ThrProThrPro----- 79
QY 303 ATCTGACTTGGAGATGTTGTGGTGAACAACATTTCTTCCCGAATCCACTTCATCC 244
DB 80 -----ProLeuGlyProHisSerProProProThrLeuSerProSerProThrPro 97
QY 243 TCTGCAATGCCACTTTGATGAGCTCAGACGACTTCTTCCAGCAGCCCAAGGCC 184
DB 98 ProPro-----ProGlyProHisSerPro 105
QY 183 GCCACGACTTG-----CAGCTCCCTGAGCTCTCTGACACGACGACGACCT 133
DB 106 ProProProLeuSerProSerProThrPro---ThrProProLeuGlyProHisSerPro 124
QY 132 CCGGAGAAAGACTTCTCATTCACACGACGACCTGGCTTCCACGAGGACAGCTGGGCT 73
DB 125 ProProThrLeuSerProSerPro---ThrProThrProProGlyProHisSerPro 143
QY 72 CCTTCTCCGGAATTCGACACGACGACCTCTCAACTCGGCTTGGGACTCGGATCC 13
DB 144 ProProPro---LeuSerProSerProThrProThrProProLeuGlyProHisSerPro 162
QY 12 CCG 10
DB 163 Pro 163

RESULT 8
US-11-036-256-31
; Sequence 31, Application US/11036256
; Publication No. US20060026719A1
; GENERAL INFORMATION:
; APPLICANT: KIELISZEWSKI, MARCIA
; APPLICANT: XU, JIANFENG
; TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND
; FILE REFERENCE: 27211/04130
; CURRENT APPLICATION NUMBER: US/11/036,256
; PRIOR FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: 60/602,562
; PRIOR FILING DATE: 2004-08-18
; PRIOR APPLICATION NUMBER: 60/582,027
; PRIOR FILING DATE: 2004-06-22
; PRIOR APPLICATION NUMBER: 60/536,486
; PRIOR FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 31
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid construct
US-11-036-256-31

Alignment Scores:
Pred. No.: 0.00755      Length: 228
Score: 123.00          Matches: 44
Percent Similarity: 34.8%      Conservative: 5
Best Local Similarity: 31.2%    Mismatches: 52
Query Match: 13.3%           Indels: 40
DB: 7                     Gaps: 7

US-10-071-510A-16 (1-493) x US-11-036-256-31 (1-228)
```

Qy	411	CCSCCTACCCCACTCTTGAACCTGGGGCCCTGG-----CCCTTTCTCCCAAGGCC	364
Db	56	ProThrProThrProProProGlyProHisSerProProProProLeuSerProSerPro	75
Qy	363	TGCTAGACAGACTACAGCGCTGCAGCATGCTGGGATAGATCCATGGGATTGATGAA	304
Db	76	ThrProThrPro-----	79
Qy	303	ATCCTGACTTTGGGATGTTGTGGTGAACCATTTTCTCCCGAATCCACTTCCATTC	244
Db	80	-----ProLeuGlyProHisSerProProProThrLeuSerProSerProThrProThr	97
Qy	243	TCTGCAGATGCAGATTCTGTATGAGAGCTCAGCAGACTTTCTCCAGCAGCTCAAGGCC	184
Db	98	ProPro-----ProGlyProHisSerPro	105
Qy	183	GCCACGACTCTG-----CGAGCTCCCTGAGTCTCTCTGCAACAGGACACCT	133
Db	106	ProProProLeuSerProSerProThrPro--ThrProProLeuGlyProHisSerPro	124
Qy	132	CGGAGAAAGACTCTGCATCACCAGCAGACGCTGGCTTCACACAGGAGAGCTGGGCT	73
Db	125	ProThrProThrLeuSerProSerPro--ThrProThrProProProGlyProHisSerPro	143
Qy	72	CCTTCTCCGGGATCTGCGCACCAGGCTCTCAACTCGAGCTCTTGGGACTCGGCAATCC	13
Db	144	ProProPro--LeuSerProSerProThrProThrProProLeuGlyProHisSerPro	162
Qy	12	CCG	10
Db	163	Pro	163

QY 87 GGGAGAGCTGGGCTCTTCT-----CCGGGAATTCGTGCACCACTCTCAAACTCGG 34  
Db 95 ProSerSerProProProAlaSerProProProSerSerProProProSerProPro 114  
QY 33 CCTCTTGGAGCTCGGATCCCGCGGCGCG 1  
Db 115 ProPheSerPro-----ProProAlaThrPro 123

RESULT 11  
US-11-188-298-694  
; Sequence 694, Application US/11188298  
; Publication No. US20060075522A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B  
; CURRENT APPLICATION NUMBER: US/11/188,298  
; PRIOR FILING DATE: 2005-07-22  
; PRIOR APPLICATION NUMBER: 60/592,978  
; PRIOR FILING DATE: 2004-07-31  
; NUMBER OF SEQ ID NOS: 22569  
; SEQ ID NO 694  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; NAME/KEY: unsure  
; LOCATION: (1)..(243)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-11-188-298-694

Alignment Scores:  
Pred. No.: 0.00837 Length: 243  
Score: 122.50 Matches: 41  
Percent Similarity: 31.7% Conservative: 36  
Best Local Similarity: 29.5% Mismatches: 26  
Query Match: 13.3% Indels: 69  
Gaps: 7

US-10-071-510A-16 (1-493) x US-11-188-298-694 (1-243)  
QY 417 TGAAGACCCCTACCCACCTCTTGACCTGGGGCTGACCCTTCTCCACAGCCCTGCTAG 358  
Db 172 TTPProProLeuProThrPhe-----LeuPheProThrPro----- 183  
QY 357 ACAGACTCAAGCGCTGACGATGCTGGAAATAGATCATGGATTGATGAGAAATCTG 298  
Db 183 ----- 183  
QY 297 ACTTTGGAGTGTGTGTGTAACATTTTCTTCCCGAATCCACTTCATCTCTGCA 238  
Db 184 -----ProLySProSerProAsnProPhe----- 191  
QY 237 GATGCCAGTTTCTGATGAGGCTCAGCAGACTTTCTTCCAAGCTCAAGGCCCGCCAG 178  
Db 192 -----PheLeuProLeu-----LeuThr 197  
QY 177 ACTTGCAGCTCCTCTGAGCTCTCTGACCAAGGACAGCCCTCGGAGAACTTCT 118  
Db 198 HLeuLeuProProPro---ProProProProProProTyHisProProPheProThrAsn 216  
QY 117 CCATCAACCAAGCCCTGCGCTTCCACCAAGGACAGCTGGGCTCTTCTCCGGGAATT 58  
Db 217 ProProProSerLeuProProProProPro-----ThrProProSerProHis 233  
QY 57 CTGCCACCAAGCCCTCAAACTGGGCTCTTGGGACTCGGATCCCGGCGGCGCG 1  
Db 234 ProProPro-----ProProProGlnPro 241

RESULT 12  
US-11-150-845-12  
; Sequence 12, Application US/11150845

/ Publication No. US20060003399A1  
; GENERAL INFORMATION:  
; APPLICANT: CytoKinectic, Inc.  
; APPLICANT: Tomasevic, Nenad  
; APPLICANT: Jia, Zhiheng  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Pierce, Daniel  
; APPLICANT: Finer, Jeffrey  
; TITLE OF INVENTION: HIGH THROUGHPUT ACTIN POLYMERIZATION ASSAY  
; FILE REFERENCE: 020552-007720US  
; CURRENT APPLICATION NUMBER: US/11/150,845  
; PRIOR FILING DATE: 2005-06-10  
; PRIOR APPLICATION NUMBER: US 60/673,444  
; PRIOR FILING DATE: 2005-04-20  
; PRIOR APPLICATION NUMBER: 60/578,949  
; PRIOR FILING DATE: 2004-06-10  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 12  
; LENGTH: 408  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-150-845-12

Alignment Scores:  
Pred. No.: 0.0129 Length: 408  
Score: 120.50 Matches: 47  
Percent Similarity: 33.9% Conservative: 9  
Best Local Similarity: 28.5% Mismatches: 50  
Query Match: 13.1% Indels: 59  
Gaps: 7

US-10-071-510A-16 (1-493) x US-11-150-845-12 (1-408)  
QY 467 CCAAGTCTTCAGAGCAGACATGTTGCCCCAGACAGGAGGGCTGTGCTGGACACCC 408  
Db 182 ProProProProSerSerArg-----GlyGlyProProProProPro 194  
QY 407 TACCCAGCTTGGACCTGGGGCTGGCCCTTCTCCACAGCCCTGATGACAGACTAC 348  
Db 195 ProProProProHisAsnSerGlyProProPro-----ProProAlaArgGlyArgGly 211  
QY 347 GCGTGCAGAGCTCTGGGAATAGATCCATGCGATTGATGAGAAATCTGACTTTGGAT 288  
Db 212 AlaProProProProProSerArg--Ala----- 220  
QY 287 GTTGTGGTAAACCATTTTCTTCCCGAATCCATTCATCTCTGCAATGCCAGTT 228  
Db 221 -----ProThrAlaAlaProProPro----- 227  
QY 227 TCTGATGAGGCTCAGACAGACTTCTTCCAGAGCTCAAGGCCGCGCAGACTGCGAG 168  
Db 228 -----ProProProSerArgProSerValAlaValPro- 238  
QY 167 CTCCTGAGCTCTCTCTGACCAAGCAGACACCTTCGAGAAAGACTTCATCAGCAG 108  
Db 239 -----ProProProProAsnArgMetGlyProPro-----ProProProA 252  
QY 107 CCAGCCCTGCGCTTCCACCAAGGAGACAGCTGGGCTCTTCTCCGGGAATTCTG----- 55  
Db 252 lAlaLeuProSerSerValProProProProProProProProProProProProProValG 272  
QY 54 -----CCACAGGCTCAAACTGGGCTCTTGGGACTCGGATCCCG 12  
Db 272 lProValAlaProProProProProProProProProProProProProGlyProProProPro 292  
QY 11 CGGGCGGCGCG 1  
Db 292 roGlyLeuPro 295

RESULT 13  
US-11-150-487-12  
; Sequence 12, Application US/11150487



Publication NO. US20060024786A1  
 GENERAL INFORMATION:  
 APPLICANT: CytoGenetics, Inc.  
 APPLICANT: Tomasevic, Nenad  
 APPLICANT: Rubselli, Alan  
 APPLICANT: Wang, Maming  
 APPLICANT: Sakowicz, Roman  
 APPLICANT: Berard, Christophe  
 TITLE OF INVENTION: WASP AND N-WASP CONSTRUCTS AND METHODS OF EXPRESSING SUCH  
 TITLE OF INVENTION: CONSTRUCTS  
 FILE REFERENCE: 020552-007910US  
 CURRENT APPLICATION NUMBER: US/11/150,487  
 PRIOR FILING DATE: 2005-06-10  
 PRIOR APPLICATION NUMBER: US 60/578,913  
 PRIOR FILING DATE: 2004-06-10  
 NUMBER OF SEQ ID NOS: 70  
 SOFTWARE: PatentIn version 3.3  
 SEQ ID NO 12  
 LENGTH: 408  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: (1)..(408)  
 OTHER INFORMATION: 98N-WASP  
 US-11-150-487-12

Alignment Scores:  
 Pred. No.: 0.0129 Length: 408  
 Score: 120.50 Matches: 47  
 Percent Similarity: 33.9% Conservative: 9  
 Best Local Similarity: 28.5% Mismatches: 50  
 Query Match: 13.1% Indels: 59  
 DB: 7 Gaps: 7

US-10-071-510a-16 (1-493) x US-11-150-487-12 (1-408)

QY 467 CCAAGTCTCAGACGACGATGTCGCCAGAGGAGGCGCTGTGAGAGACCC 408  
 DB 182 ProProProProSerArg-----GlyGlyProProProProPro 194  
 QY 407 TACCCACCTTGTACCTGGGCGCTGCGCTTCTCCACAGCCCTGTAGACAGCTCAC 348  
 DB 195 ProProProHisAsnSerGlyProProPro-----ProProAlaArgGlyArgGly 211  
 QY 347 GCGTCGACGATGCTGGGAATAGATCCATGAGATTGATGAGAAATCTGACTTTGGGAT 288  
 DB 212 AlaProProProProProSerArg--Ala----- 220  
 QY 287 GTTGTGTGTAACCATTTTCTCCCGGATCCATTCATCTCTGACAGATGCCAGTT 228  
 DB 221 -----ProThrAlaAlaProProPro----- 227  
 QY 227 TCTGATGAGCTCAGACACTTTCTTCAGACAGCTCAAGCGCGCAGCACTCTGCCAG 168  
 DB 228 -----ProProProSerArgProSerValAlaValPro- 238  
 QY 167 CTCCTGAGCTCTCTGACACGAGCAGACCTCCGGAGAAACATTCATCAGACAG 108  
 DB 239 -----ProProProProAsnArgMetTyrProPro-----ProProProA 252  
 QY 107 CCAGCCCTGGGCTTCCACGAGGACAGCTGGGCTCTCTCCGGGAATTCG----- 55  
 DB 252 lAlauProSerSerAlaProSerGlyProProProProProProSerValIleuGlyValG 272  
 QY 54 -----CCACGAGCTCTCAAACTCGGCTCTTGGGAATCGGATCCCC 12  
 DB 272 lProValAlaProProProProProProProProProProGlyProProProProP 292  
 QY 11 CGGGCCGGCGG 1  
 DB 292 roGlyLeuPro 295

RESULT 14

US-11-134-563-6  
 Sequence 6, Application US/11134563  
 Publication No. US20050287569A1

GENERAL INFORMATION:  
 APPLICANT: Leon, John M.  
 APPLICANT: Campellone, Kenneth G.  
 TITLE OF INVENTION: ESPRU NUCLEIC ACIDS AND PROTEINS AND  
 TITLE OF INVENTION: USES THEREOF  
 FILE REFERENCE: 07917-280001  
 CURRENT APPLICATION NUMBER: US/11/134,563  
 PRIOR FILING DATE: 2005-05-20  
 PRIOR APPLICATION NUMBER: US 60/573,600  
 PRIOR FILING DATE: 2004-05-20  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 6  
 LENGTH: 505  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-11-134-563-6

Alignment Scores:  
 Pred. No.: 0.0131 Length: 505  
 Score: 120.50 Matches: 47  
 Percent Similarity: 33.9% Conservative: 9  
 Best Local Similarity: 28.5% Mismatches: 50  
 Query Match: 13.1% Indels: 59  
 DB: 7 Gaps: 7

US-10-071-510a-16 (1-493) x US-11-134-563-6 (1-505)

QY 467 CCAAGTCTCAGACGACGATGTCGCCAGAGGAGGCGCTGTGAGAGACCC 408  
 DB 279 ProProProProSerArg-----GlyGlyProProProProPro 291  
 QY 407 TACCCACCTTGTACCTGGGCGCTGCGCTTCTCCACAGCCCTGTAGACAGCTCAC 348  
 DB 292 ProProProHisAsnSerGlyProProPro-----ProProAlaArgGlyArgGly 308  
 QY 347 GCGTCGACGATGCTGGGAATAGATCCATGAGATTGATGAGAAATCTGACTTTGGGAT 288  
 DB 309 AlaProProProProProSerArg--Ala----- 317  
 QY 287 GTTGTGTGTAACCATTTTCTCCCGGATCCATTCATCTCTGACAGATGCCAGTT 228  
 DB 318 -----ProThrAlaAlaProProPro----- 324  
 QY 227 TCTGATGAGCTCAGACACTTTCTTCAGACAGCTCAAGCGCGCAGCACTCTGCCAG 168  
 DB 325 -----ProProProSerArgProSerValAlaValPro- 335  
 QY 167 CTCCTGAGCTCTCTGACACGAGCAGACCTCCGGAGAAACATTCATCAGACAG 108  
 DB 336 -----ProProProProAsnArgMetTyrProPro-----ProProProA 349  
 QY 107 CCAGCCCTGGGCTTCCACGAGGACAGCTGGGCTCTCTCCGGGAATTCG----- 55  
 DB 349 lAlauProSerSerAlaProSerGlyProProProProProProSerValIleuGlyValG 369  
 QY 54 -----CCACGAGCTCTCAAACTCGGCTCTTGGGAATCGGATCCCC 12  
 DB 369 lProValAlaProProProProProProProProProProGlyProProProProP 389  
 QY 11 CGGGCCGGCGG 1  
 DB 389 roGlyLeuPro 392

RESULT 15  
 US-11-150-845-4  
 Sequence 4, Application US/11150845  
 Publication No. US20060003399A1  
 GENERAL INFORMATION:

Search completed: April 29, 2006, 04:09:46  
Job time : 23 secs

APPLICANT: Cycokinetics, Inc.  
APPLICANT: Tomasevic, Nenad  
APPLICANT: Jia, Zhiheng  
APPLICANT: Sakowicz, Roman  
APPLICANT: Pierce, Daniel  
APPLICANT: Finer, Jeffrey  
TITLE OF INVENTION: HIGH THROUGHPUT ACTIN POLYMERIZATION ASSAY  
FILE REFERENCE: 020552-007720US  
CURRENT APPLICATION NUMBER: US/11/150,845  
PRIOR FILING DATE: 2005-06-10  
PRIOR APPLICATION NUMBER: US 60/673,444  
PRIOR FILING DATE: 2005-04-20  
PRIOR APPLICATION NUMBER: 60/578,949  
PRIOR FILING DATE: 2004-06-10  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 4  
LENGTH: 505  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(505)  
OTHER INFORMATION: FL-N-WASP  
US-11-150-845-4

Alignment Scores:  
Pred. No.: 0.0131 Length: 505  
Score: 120.50 Matches: 47  
Percent Similarity: 33.9% Conservative: 9  
Best Local Similarity: 28.5% Mismatches: 50  
Query Match: 13.1% Indels: 59  
DB: 7 Gaps: 7

US-10-071-510a-16 (1-493) x US-11-150-845-4 (1-505)

QY 467 CCNAGTCTCAGACGAGCATGTTGCCCGACAGAGGAGGCGCTGTGCTGGAGACCC 408  
DB 279 ProProProProSerArg-----GlyGlyProProProProPro 291  
QY 407 TACCCACTTGTGACCTGGGCGCTGGCCCTTCTCCACAGCCCTGTAGACAGACTCAG 348  
DB 292 ProProProHisbSerglyProProPro-----ProProHisArgGlyArgGly 308  
QY 347 GCGTCACGATGCTGGGATAGATCCATGGGATTGATGAGAAATCTGACTTTGGGAT 288  
DB 309 AlaProProProProSerArg--Ala----- 317  
QY 287 GTTGTGTGMAAACCATTTTCTCCCGAATCCACTTCATCTCTGCGATGCCAGTT 228  
DB 318 -----ProthAlaAlaProProPro----- 324  
QY 227 TCTGATGAGGCTCAGCAGACTTCTTCCAGACGCTCAAGGCCCGCCAGACTGTCCAG 168  
DB 325 -----ProProProSerArgProSerValAlaValPro- 335  
QY 167 CTCCTGAGACTCTCTCTGACACACGCGGACGACCTCCGAGAGAACTTCTCCATCACAG 108  
DB 336 -----ProProProProAsnArgMetGlyProPro-----ProProProA 349  
QY 107 CCAGCCCTGGGCTTCCACGAGGAGAGCTGGGCTCTTCTCCGGGAAATTTCTG----- 55  
DB 349 lAlaLeuProSerSerAlaProSerGlyProProProProProProSerValLeuGlyValG 369  
QY 54 -----CCACGAGCTTCAAACTCGGCTCTTGGGAGCTGGGATCCCC 12  
DB 369 lYProValAlaProProProProProProProProProProProProProProProPro 389  
QY 11 CCGGCGCGCGCG 1  
DB 389 roGlyLeuPro 392